Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder

Prevalence and correlates of cognitive impairment in euthymic adults with bipolar disorder: a systematic review

Breda Cullen[1]*, Joey Ward[1], Nicholas A. Graham[1], Ian J. Deary[2], Jill P. Pell[1], Daniel J. Smith[1] and Jonathan J. Evans[1]

1. Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
2. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

*Correspondence:

Dr Breda Cullen

R4, University Section

Ground Floor, Office Block

Queen Elizabeth University Hospital

Glasgow G51 4TF

Tel.: +44 141 451 5879

Email: breda.cullen@glasgow.ac.uk

Word count 5,708 (excluding references)

ABSTRACT

Background

Previous reviews have identified medium-large group differences in cognitive performance in adults with bipolar disorder (BD) compared to healthy peers, but the proportion with clinically relevant cognitive impairment has not yet been established. This review aimed to quantify the prevalence of cognitive impairment in euthymic adults with BD, and to describe sociodemographic, clinical and other factors that are significantly associated with cognitive impairment.

Methods

Systematic literature review. The population was euthymic community-dwelling adults with BD, aged 18-70 years, and recruited consecutively or randomly. The outcome was cognitive impairment, relative to healthy population norms. Electronic databases and reference lists of relevant articles were searched, and authors were contacted. Original cross-sectional studies published in peer-reviewed English-language journals from January 1994 to February 2015 were included. Methodological bias and reporting bias were assessed using standard tools. A narrative synthesis is presented together with tables and forest plots.

Results

Thirty articles were included, of which 15 contributed prevalence data. At the 5^{th} percentile impairment threshold, prevalence ranges were: executive function 5.3% to 57.7%; attention/working memory 9.6% to 51.9%; speed/reaction time 23.3% to 44.2%; verbal memory 8.2% to 42.1%; visual memory 11.5% to 32.9%. More severe or longstanding illness and antipsychotic medication were associated with greater cognitive impairment.

Limitations

The synthesis was limited by heterogeneity in cognitive measures and impairment thresholds, precluding meta-analysis.

Conclusions

Cognitive impairment affects a substantial proportion of euthymic adults with BD. Future research with more consistent measurement and reporting will facilitate an improved understanding of cognitive impairment burden in BD.

Keywords Bipolar disorder; cognition; neuropsychology; impairment; prevalence

Registration number PROSPERO reference number CRD42015017558

INTRODUCTION

Bipolar disorder (BD) is known to be associated with cognitive impairment, which persists between illness episodes and contributes to functional disability. Impairment is typically found on tests of attention, working and episodic memory, processing speed and executive function, with significant group differences of medium to large effect size compared to healthy comparison groups (Arts et al., 2008; Bourne et al., 2013; Mann-Wrobel et al., 2011; Robinson et al., 2006). Although such group-level differences have been consistently reported, the proportion of adults with BD who have clinically relevant levels of cognitive impairment has not yet been clearly established. It is likely that there is marked within-group variation, ranging from normal performance through to severe multi-domain impairment. It has been argued that if overall group differences are being driven by a subgroup of patients with marked levels of impairment, this serves to obscure the true picture of cognitive impairment in the BD population, which in fact may be severe for some and absent for many others (Iverson et al., 2011).

There are a number of reasons why it would be beneficial to establish the prevalence of cognitive impairment in the BD population. From a clinical point of view, cognitive impairment is a major contributor to the overall burden of disability in mood disorders, and is a target in its own right for therapeutic intervention. Service planning would be helped by clearer information about numbers and characteristics of those who are likely to need more clinical or social care input to manage the disabling effects of cognitive impairment. From a research perspective, shifting our focus to identifying subgroups with cognitive impairment will facilitate efforts to understand why some people with BD experience significant problems with cognition while others remain unimpaired. This, in turn, may help to identify particular risk factors for clinically significant cognitive impairment.

Objectives

1. To determine the prevalence of cognitive impairment in euthymic adults with a history of BD.
2. To describe sociodemographic, clinical and other factors that are associated with cognitive impairment in BD.

Scope of review

The population of interest was community-dwelling adults with a history of BD (the exposure), who were euthymic at the time of assessment. The outcome of interest was cognitive impairment, measured using standardised tests; presence or absence of impairment was defined with reference to healthy population norms. Since the aim was to determine prevalence, only cross-sectional results were considered (cross-sectional studies or baseline results from cohort studies or trials).

MATERIALS AND METHODS

The review was conducted according to a structured protocol which followed PRISMA-P guidance (Moher et al., 2015). The protocol was registered on the PROSPERO database on 16 March 2015 (reference number CRD42015017558; http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015017558). Reporting is in accordance with PRISMA and MOOSE guidelines (Moher et al., 2009; Stroup et al., 2000).

Eligibility criteria

The following inclusion criteria were applied during the search and screening process: original research published in peer-reviewed journals from 1994 onwards (the year that DSM-IV and ICD-10 diagnostic classifications came into use); articles published in English; studies of community-dwelling adults (not hospital in-patients) aged 18 to 70 years inclusive (to minimise the additional contribution of age-related cognitive decline); cross-sectional studies or baseline results from cohort studies or trials; clinical samples must have been recruited consecutively from clinics or via a method

that ensured eligible individuals in the target population had an equal chance of being approached (so that prevalence estimates would be based on representative samples); primary diagnosis of BD; euthymic at time of assessment; assessed using at least one direct, standardised, objective cognitive measure. Articles were excluded if samples were selected on the basis of presence/absence of cognitive impairment (known or suspected).

Concepts and definitions

Bipolar disorder

History of bipolar disorder type I, II or not otherwise specified, meeting defined criteria (e.g. DSM or ICD).

Euthymia

Not meeting defined criteria for a depressive or manic episode at time of cognitive assessment; or as otherwise defined by the study authors based on an appropriate clinical measure.

Cognitive impairment

Evidence of impaired performance on one or more objective cognitive tests. Impairment was defined as the fail range on a pass/fail test, or as otherwise defined by the study authors with reference to the score distribution of a healthy comparison group (e.g. from published test norms, or an appropriate comparison group recruited to the study). Results based on any threshold that was less strict than 1 standard deviation (SD) below the comparison mean would not be considered.

Prevalence

Assessments must have been conducted at a single time point, yielding a point prevalence estimate of cognitive impairment, reported as the proportion of the sample falling below the cut-off for impairment.

Correlates

Any sociodemographic, clinical or other factor that was reported by the authors to be significantly associated with presence or severity of cognitive impairment.

Search strategy

Information sources

The following electronic databases were searched on 24 February 2015: Web of Science (Thomson Reuters), including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Current Contents Connect, Data Citation Index, MEDLINE and SciELO Citation Index; PubMed (NCBI), including MEDLINE, PubMed Central and in-process/ahead-of-print citations; EBSCOhost (EBSCO), including CINAHL and PsycINFO. Additional articles published up to the search date were sought via: the 'cited by' function within individual electronic records of relevant articles; hand searching of reference lists of relevant articles and recent review papers; and email contact with authors.

Process for study identification and selection

A detailed search strategy was developed and tailored for each electronic database. Controlled vocabulary and free text variations were used, including synonyms, abbreviations and spelling variants. Appendix 1 shows the search strategy as implemented in Web of Science. Search outputs were managed using EndNote software.

Duplicate records were removed, and study titles and/or abstracts were screened for relevance by B.C. Screening was carried out with reference to a detailed checklist of eligibility criteria; this was piloted by B.C. and J.W. independently against a sample of initial search results, and refined as required (see Appendix 2 for the final checklist). The sensitivity of the search strategy was checked by testing whether key papers that were known to be relevant were detected by the search. Reproducibility was assessed by J.W., who independently ran the search in one electronic database (Web of Science) and screened the first 200 titles and/or abstracts for relevance. Agreement was 93% (100% following consensus discussion). Full text was obtained for all potentially relevant papers that remained. These were assessed by B.C. using the eligibility checklist, with J.W. independently

assessing the first 100 papers for comparison. Agreement was 95% (100% following consensus discussion). Reasons for exclusion were documented.

Data extraction

A spreadsheet template was used for extracting data from included papers, having been piloted by B.C. and J.W. independently. The list of data fields is given in Appendix 3. Data extraction was carried out by B.C., following which J.W. compared four randomly-chosen data extraction records against the source papers to check for accuracy and completeness; no discrepancies were identified. Where authors appeared to have collected data that could be used to report prevalence of impairment, but had not reported prevalence explicitly in the paper (e.g. articles only reporting group mean differences), the authors were contacted via email to request prevalence results using an appropriate cut-off of their choice.

Assessment of risk of bias

Risk of bias within studies

Each included study was assessed for risk of bias using a critical appraisal tool for systematic reviews addressing questions of prevalence (Munn et al., 2014). Reporting bias was assessed using the STROBE checklist for cross-sectional studies (von Elm et al., 2007). B.C. and N.A.G. independently rated randomly-chosen articles for comparison, followed by consensus discussion. Initial rating concordance was 83-95% across four articles, and 93% when one further article was independently assessed following the consensus discussion exercise. Subsequent ratings were made by B.C. only. These assessments were considered in the synthesis and discussion, in order to comment on the quality of the literature in this field and to aid interpretation of the results.

Risk of bias across studies

Funnel plots were generated using the metafunnel package in Stata v13 (College Station, TX: StataCorp LP), to allow visual inspection of the relationship between magnitude and precision of prevalence estimates. These are scatter plots depicting a measure of study size (e.g. sample size or standard error of the effect estimate) on the vertical axis against the study's effect estimate on the horizontal axis. Larger (more precise) studies are expected to have effect estimates close to the centre on the horizontal axis, and smaller studies are expected to have effect estimates scattered symmetrically about the centre. Asymmetry in this characteristic inverted funnel shape indicates "small study bias", for example resulting from publication bias (Egger et al., 1997).

Data synthesis

Where one study population was analysed in two or more eligible articles, the article reporting the largest sample was included in the data synthesis. Additional articles were only included if they contributed unique relevant information (e.g. additional cognitive measures). A narrative synthesis is presented, alongside summary tables of extracted data (Tables 1 to 3 and Supplementary Tables S1 to S3), and forest plots of impairment prevalence estimates and 95% confidence intervals (CI) by cognitive domain (Figures 2 to 5). Forest plots were generated using the metan package in Stata v13 (College Station, TX: StataCorp LP). Sociodemographic, clinical and other variables that were significantly associated with cognitive impairment were summarised, and consistency in these findings was compared across studies (Supplementary Figure S4). Only variables that were potential risk factors for impairment were included; variables that were viewed as consequences of that impairment (e.g. occupational status, instrumental functioning) were not considered, on the basis that they are not potential causal, mediating or moderating factors in explaining the association between BD status and cognitive impairment.

RESULTS

Article selection

Figure 1 shows a PRISMA flow diagram of the article selection process. Titles and/or abstracts of 5,412 records were screened for eligibility, followed by full text evaluation of 658 papers. Forty-six articles were deemed eligible. The most common reasons for exclusion were lack of evidence of consecutive sample recruitment, inclusion of in-patients in study samples, and inclusion of non-euthymic participants. Examples of acceptable sample recruitment methods in the eligible articles were systematic invitation of: consecutively attending eligible patients at out-patient clinics; all eligible patients on a database of open records at a specific clinical service; all eligible persons identified via national registers during a specific period. Of the 46 eligible articles, 16 were omitted from the data synthesis (see list in Appendix 4): 11 reported on overlapping samples without contributing relevant additional information, and for a further five, results directly addressing the two research questions of this review were unavailable.

Characteristics of included studies

Key characteristics of the 30 included articles are summarised in Table 1. The majority included BD-I samples only (Altshuler et al., 2004; Arslan et al., 2014; Cavanagh et al., 2002; Cheung et al., 2013; Doganavsargil-Baysal et al., 2013; Fakhry et al., 2013; Ferrier et al., 1999; Frangou et al., 2005; Goswami et al., 2009; Ibrahim et al., 2009; Jamrozinski et al., 2009; Juselius et al., 2009; Kieseppa et al., 2005; Lopera-Vasquez et al., 2011; Lopez-Jaramillo et al., 2010; Normala et al., 2010; Osher et al., 2011; Pirkola et al., 2005). A further eight articles reported on mixed BD samples (Barrera et al., 2013; Daban et al., 2012; Elshahawi et al., 2011; Martino et al., 2014; Martino et al., 2008; Mur et al., 2007; Sánchez-Morla et al., 2009; van der Werf-Eldering et al., 2010) and four articles included separate BD-I and BD-II samples (Martino et al., 2011a; Martino et al., 2011b; Martino et al., 2011c; Sparding et al., 2015). Three articles reported on samples recruited from population registers of twin births and hospital discharges (Juselius et al., 2009; Kieseppa et al., 2005; Pirkola et al., 2005) and the rest recruited from specialist psychiatry clinics. Definitions of euthymia differed across studies; many used the Hamilton Rating Scale for Depression (HRSD) and the Young Mania Rating Scale

(YMRS), but score thresholds varied. Most studies excluded participants with major psychiatric, neurological or medical comorbidity or learning disability, and many also excluded those with recent substance misuse or electroconvulsive therapy.

Ratings of methodological and reporting bias are shown in Supplementary Figures S1 and S2, respectively. Although all studies aimed to recruit representative participants using consecutive or random methods, nine of 30 articles included samples which were unrepresentative of the BD population with regard to gender balance and two did not report gender composition. Most articles did not report numbers of patients initially considered or deemed eligible, or information about comparability of eligible patients who did and did not participate; there was evidence of adequate coverage of the intended population in only four articles. Sample sizes were generally small, with only seven studies having 50 or more per group. All articles reported on objective cognitive measures, but 13 did not report sufficient information to allow appraisal of measurement reliability (e.g. qualifications and training of assessors; inter-rater reliability data). Most did not report adequate consideration of sources of bias or imprecision in their procedures or interpretation.

Prevalence of cognitive impairment

Prevalence was available for 15 articles, reporting on 16 BD samples. Tables 2 and 3 show prevalence results in BD-I only and mixed BD samples, respectively. Characteristics of these samples are provided in Supplementary Tables S1 and S2. Prevalence was available for one study with separate BD-I and BD-II samples (Supplementary Table S3).

Studies applied a variety of impairment thresholds: some were simple pass/fail cut-offs, and others were based on score distributions from published test norms or from a healthy comparison group. Distribution-based thresholds ranged from 1SD to 2SD below comparison mean, with the most common being 1.5SD (approximately 7th percentile), 1.64SD (approximately 5th percentile) and 2SD (approximately 2nd percentile). At every threshold and on almost all cognitive measures, prevalence of impairment in BD samples was higher than in the comparison group. Heterogeneity in

prevalence across studies did not clearly relate to study quality/risk of bias. Studies differed in whether they used comparison group score distributions or published norms as the reference for impairment, but there was no clear relationship between choice of reference and magnitude of impairment prevalence. For example, on the same tests at the same thresholds, Mur et al. (2007) used published norms and reported lower prevalence estimates than Juselius et al. (2009), who used their own comparison group. On the other hand, Cheung et al. (2013) used published norms and reported some of the highest prevalence estimates across several cognitive domains. Prevalence of impairment did not differ consistently between BD-I only (Table 2) and mixed BD samples (Table 3), although direct comparison is difficult owing to the variation in measures and thresholds used. In the only study where BD-I and BD-II samples could be directly compared (Sparding et al., 2015) (Table S3), prevalence was higher in the BD-I participants on several measures, but there was considerable overlap between the two samples.

Prevalence of cognitive impairment was further considered according to cognitive domain. Results within domains are presented graphically using forest plots, but pooled estimates are not reported because of the wide variation in cognitive tests used and in cut-offs applied to define presence of impairment. The classification of tests by domain was guided by the classifications used by the authors of the original articles. Where tests were thought to cross multiple domains, this is indicated in Tables 2 and 3.

Executive function, reasoning and social cognition

Figure 2 shows prevalence of impairment in studies which used a normative distribution-based threshold for impairment. Additional score-based threshold results from three studies (Altshuler et al., 2004; Barrera et al., 2013; Normala et al., 2010) are reported in Tables 2 and 3. Measures which are significantly influenced by performance speed are considered separately from those that are not, to minimise the overlap between underlying processing speed ability and instrumental executive function. The former category included timed fluency measures, Stroop test, Trailmaking test, and

composite scores primarily influenced by these. The latter category included Tower tests, non-time-dependent aspects of fluency tasks (e.g. category switching accuracy), reasoning tests, Wisconsin Card Sorting test, BADS Six Elements task, and composite scores primarily influenced by these.

Figure 2 shows that impairment prevalence tended to be slightly higher on speed-sensitive tasks (panel B) than on those that depend less on speed (panel A), though this pattern was not evident in all studies. The estimates did not follow a clear gradient according to the different threshold strata: for example, the estimates from Cavanagh et al. (2002) were the same at the 5th and 2nd percentile levels. This may be a consequence of small sample size, or may indicate that impaired individuals were strongly clustered at the extreme low end of the score distribution, such that less strict thresholds made little difference to the absolute numbers considered impaired. The estimates from Juselius et al. (2009) were somewhat higher than expected in the context of the other studies. This may be related to study size and quality, but it should also be noted that this study included several twin pairs who were concordant for BD. BD-II-only results are not shown in Figure 2, but Supplementary Table S3 indicates that fewer BD-II participants were impaired, in comparison with BD-I participants, on most executive function measures in Sparding et al. (2015). Only one study provided prevalence data for social cognition tasks (Barrera et al., 2013): in a small mixed BD sample (n = 12), prevalence of impairment on emotional and cognitive theory of mind measures was higher compared to the healthy comparison sample (Table 3).

Attention and working memory

Figure 3 shows the prevalence of impairment in five studies (of similar quality) that reported attention/working memory measures. Estimates were generally higher than in the healthy comparison population, and this was most striking on the CNS-VS complex attention score reported by Cheung et al. (2013) and the Mindstreams attention score from Osher et al. (2011). These scores are composites of several demanding tasks, more akin to the executive measures presented in Figure 2. Additional measures from Normala et al. (2010) are reported in Table 2, showing a slightly

elevated percentage of BD participants with reduced forward and backward digit span. The study by Sparding et al. (2015) allows comparison between BD-I and BD-II samples on two attention/working memory measures, indicating that proportions impaired were similar (Supplementary Table S3).

Speed and reaction time

Figure 4 shows that prevalence of impairment on speed and reaction time measures was similar across different impairment thresholds. However, Daban et al. (2012) reported that 30.2% were impaired on the WAIS-III Digit Symbol Coding task at the 5th percentile threshold, whereas Sparding et al. (2015) reported 19% impairment prevalence on the same task at the less strict threshold of 11th percentile. Daban et al. assessed a mixed BD sample but did not report subtypes or illness characteristics, making it difficult to infer reasons for the disparity with Sparding et al.'s BD-I sample. It was also evident from the Sparding et al. study that fewer BD-II participants were impaired on these tasks; in the case of WAIS-III Digit-Symbol Coding, the proportion impaired (11%) was in line with the normative score distribution (Supplementary Table S3).

Memory

Figure 5 shows impairment prevalence results for verbal memory (panel A) and visual memory (panel B). Additional score-based threshold results from Ibrahim et al. (2009) are shown in Table 2. Two studies of similar quality that reported composite verbal and visual measures separately (Cheung et al., 2013; Sánchez-Morla et al., 2009) showed contradictory findings regarding relative prevalence of impairment: both studies reported that 28.8% were impaired on verbal memory at the 5th percentile threshold, but proportions impaired on visual memory were 11.5% in Cheung et al. (2013) versus 32.9% in Sánchez-Morla et al. (2009). The proportions impaired on overall memory composite measures were 43.1% at the 7th percentile threshold (Osher et al., 2011) and 30.8% at the 5th percentile (Cheung et al., 2013).

The California Verbal Learning Test (CVLT) was the most common of the verbal measures, used in four studies with different thresholds. Results from Cavanagh et al. (2002) and Altshuler et al. (2004) indicated a threshold-related gradient, with fewer participants falling below the stricter 2nd percentile level for CVLT learning and recall, though not for recognition performance. Sánchez-Morla et al. (2009) reported lower impairment prevalence than Cavanagh et al. using the same 5th percentile threshold for the same CVLT measures (total trials 1 to 5, and long delay recall). This may be explained by the larger sample size and mix of BD-I and BD-II participants in the former study. No verbal memory results were available for BD-II separately.

Visual memory results were available from four studies of similar quality. Three (Mur et al., 2007; Sánchez-Morla et al., 2009; Sparding et al., 2015) used the Rey Complex Figure Test (RCFT) at different impairment thresholds; prevalence on this test was lowest in Sparding et al. (2015) despite the less strict threshold and more severe clinical characteristics of their sample. Prevalence of visual memory impairment was similar between BD-I and BD-II samples in that study (Supplementary Table S3).

Visuospatial function

Three studies (Osher et al., 2011; Sánchez-Morla et al., 2009; Sparding et al., 2015) reported visuospatial measures (Tables 2, 3 and S3). Impairment prevalence was lower for visuospatial tasks than for other cognitive domains, though still somewhat higher than would be expected from the normative distribution. Prevalence was highest on the WAIS-III Block Design task—reported as 40% by Sparding et al. (2015) at the 11th percentile threshold—which may reflect the executive and speed components that contribute to success on this task. Prevalence was similarly high among BD-II participants on this task (Sparding et al., 2015).

Any domain, multi-domain and global impairment

Fakhry et al. (2013) used the Mini-mental State Examination (MMSE), Mental Test Score (MTS) and Clock Drawing Test (CDT)—typically used as global measures in dementia settings—to assess BD-I participants, grouped by the polarity of their most recent illness episode. Table 2 shows that the proportions falling below the impairment cut-off were markedly higher in the group whose most recent episode was manic. No BD participant scored below the cut-off on the CDT, however.

Osher et al. (2011) reported that 49% of their BD-I sample fell below the 7th percentile (1.5SD) on the global cognition measure of the Mindstreams computerised battery. Also in BD-I, 46.2% of the Cheung et al. (2013) sample were below the 5th percentile on the CNS-VS overall measure of neurocognition. Furthermore, 61.5% were at least 1SD below the normative mean on at least two CNS-VS index scores, and 40.4% met the stricter criterion of being at least 2SD below the normative mean on at least two index scores.

Two studies reported overall results from mixed BD samples. Van der Werf-Eldering et al. (2010) found that 6 of 46 participants (13%) were at least 2SD above the healthy comparison mean (where higher scores indicated worse performance) in at least one cognitive domain. The sample of 46 was a euthymic sub-group from a larger study, for whom demographic and clinical characteristics were not available. It is therefore unclear why the proportion impaired was relatively low in this study. Martino et al. (2014) assessed a larger sample (n = 100), and reported that 70% were impaired using “soft” criteria (1.5SD below the normative mean in at least one cognitive domain) and 30% were impaired using “hard” criteria (at least 2SD below the normative mean on at least two domains).

Risk of bias across studies

Supplementary Figure S3 shows funnel plots of the relationship between prevalence estimates and their precision (standard error), presented separately by cognitive domain, for studies reporting measures at the 5th percentile impairment threshold. Visual inspection suggested a degree of asymmetry for measures of verbal memory, and to a lesser extent for speed-sensitive measures

(both within the executive domain and on specific tests of speed/reaction time). Relatively fewer estimates in the lower left quadrant of these plots may indicate publication bias, or reflect other factors such as different sample characteristics or assessment methods in the smaller/less precise studies. The small number of independent measures meant it was not possible to apply statistical tests of asymmetry.

Factors associated with cognitive impairment

Twenty-eight articles provided information regarding the association between various sociodemographic, clinical or other variables and presence or severity of cognitive impairment. Articles were not always clear about which associations had been tested statistically, and they varied in the extent to which they adjusted for potential confounders. Supplementary Figure S4 shows an overview of the types of variables tested, with significant findings highlighted across studies.

Associations with demographic variables and premorbid ability were often not tested. In some cases this was because key background factors had been frequency-matched in a between-group study design, or had been adjusted for when calculating standardised cognitive scores. Other analyses included these background factors as covariates (e.g. in multiple regression), without reporting results for these covariates separately. For the remainder, greater cognitive impairment was associated with older age and lower education and premorbid ability in some studies, but others reported no significant findings.

Illness characteristics—such as duration since onset, number of affective episodes and hospitalisations, history of psychotic symptoms, and residual depressive or manic symptoms—were more frequently investigated. Where significant results were reported, they indicated that more severe illness characteristics were associated with worse cognitive function. An exception was history of psychotic symptoms, for which one study reported both positive and negative effects. Several studies investigated associations with psychotropic medication, with mixed findings. The strongest evidence of association was between antipsychotic medication and worse cognition,

though some studies reported null findings. By contrast, mood stabilisers (lithium or anticonvulsants) were less frequently associated with impairment.

Although two studies examined history of alcohol/substance use disorder, none investigated the relationship between frequency/amount of alcohol or recreational drug consumption and cognitive impairment. No study examined associations with smoking or other cardiovascular risk factors that may be relevant to cognitive impairment.

DISCUSSION

Summary of findings

The aims of this review were to determine the prevalence of cognitive impairment in euthymic adults with BD, and to ascertain which clinical, sociodemographic or other factors were associated with cognitive impairment in this population. Thirty articles contributed to the findings, of which 15 provided prevalence. Impairment prevalence was similar between BD-I only and mixed BD samples. One study with separate results for BD-I and BD-II participants indicated that impairment was more common in those with BD-I, though considerable overlap was apparent. Examination of impairment proportions across different cognitive domains indicated wide variation both within and between domains. For example, taking the 5th percentile threshold as the reference, impairment prevalence ranges were as follows: non-speed-sensitive executive function 5.3% to 57.7%; speed-sensitive executive function 10.0% to 36.8%; attention/working memory 9.6% to 51.9%; speed/reaction time 23.3% to 44.2%; verbal memory 8.2% to 42.1%; visual memory 11.5% to 32.9%. Generally small sample sizes resulted in wide CIs for most estimates. A recent review of neuropsychological function in BD (Szmulewicz et al., 2015) highlighted impairment prevalence as an issue of particular interest, and reported estimates between 30% and 57% from six studies. Four of these studies were not eligible for the present review, either because participants were not euthymic or because the recruitment method did not meet our criteria. The fact that the lower bounds of the prevalence estimates reported in the present review are below the previous estimate of 30% can be understood

in light of our exclusion of non-euthymic participants and samples recruited by convenience, either of which may bias prevalence estimates upwards.

There was some evidence that more severe or longstanding illness was associated with greater cognitive impairment. Several studies reported an association with antipsychotic medication but less so with other types of psychotropic medication; it should be noted, however, that medication associations are likely to be confounded by illness severity as well as treatment adherence and responsiveness. A previous individual participant data meta-analysis of 2,876 euthymic patients with BD (Bourne et al., 2013) also reported significant correlations between cognitive impairment and some illness severity indices (number of manic episodes and total hospitalisations), and reported an association for antipsychotic medication only, but not lithium, antidepressants or anticonvulsants.

Limitations of included studies

Valid prevalence estimates depend on representative samples, but representativeness was questionable in many of the studies included here. Although all appear to have used an appropriate recruitment method (e.g. consecutive or random sampling), details were scant in published papers regarding exact recruitment processes and numbers considered at each stage. Exclusion on the basis of comorbidity such as substance misuse was common, but numbers excluded were rarely reported. Definitions of euthymia varied; even when these were based on common measures (e.g. HRSD and YMRS), cut-off scores differed across studies. Some degree of residual affective symptoms was present in most BD samples, but this was not always considered as a confounding factor in analyses. A wide range of cognitive tests was used, and even within specific tests, many different scores were reported (e.g. CVLT sub-scores). This made direct comparison across studies difficult. The use of different thresholds to define cognitive impairment also limited synthesis at the outcome level. Most studies focused on the cognitive domains of executive function, memory and attention, with other areas of function such as visuospatial ability, language and praxis studied rarely if at all. Articles were

sometimes unclear regarding which demographic, clinical or other variables were statistically tested against cognitive measures.

Limitations of review

We aimed to follow best practice in systematic review methodology, but reproducibility of screening, data extraction and bias appraisal processes was checked for only a proportion of records. Judgements about study eligibility relied solely on information contained in the articles, and authors were not contacted to request missing information during the selection process. A large number of articles were excluded on the sample recruitment criterion, in some cases because this information was not contained in the article; it is possible that some of these did in fact employ an appropriate sampling procedure. The requirement for information within the article indicating an acceptable procedure meant that several articles included in previous reviews of cognition in BD are not included here, including some that reported prevalence estimates. Despite repeated attempts to obtain additional prevalence results from authors of eligible articles, prevalence data were available for only 15 articles. In particular, there was little information regarding impairment prevalence in BD-II samples. Heterogeneity of cognitive measures and thresholds meant that it was not feasible to meta-analyse the prevalence estimates obtained, or to conduct statistical tests of funnel plot asymmetry, and so the results are limited to graphical and narrative synthesis only. This was organised by cognitive domain, but we acknowledge that many tests make multiple cognitive demands across domains. Regarding our second research question, variation in the way that correlates were analysed across studies meant it was not possible to comment on the nature of any inter-relationships between the potential risk factors reported here. Risk of bias was considered carefully, but it should be noted that the appraisal tool used was developed for questions of prevalence, whereas many of the studies included here were not originally designed to investigate prevalence. The literature search results were restricted to English-language publications only, although studies from a wide range of international settings were found.

Conclusions and implications

This review is the first to systematically examine the prevalence of cognitive impairment in euthymic bipolar disorder. It complements and extends the findings of previous reviews, which have focused on magnitude of between-group differences on cognitive measures. Although group differences are important for understanding the nature and extent of cognitive impairment in this population, quantifying the number who have clinically relevant cognitive impairment is essential if we wish to identify risk factors for a cognitively impaired subtype of euthymic BD, and to target clinical resources towards neuropsychological rehabilitation and support for those who need it most. Despite the heterogeneity in the present findings, some tentative conclusions can be drawn. Cognitive impairment affects patients across the BD spectrum; impairment appears to be more common in BD-I but there is considerable overlap with BD-II. It is also evident that even at the lower ends of the prevalence ranges reported here, the proportion of patients whose affective illness is in remission but who continue to show cognitive impairment substantially exceeds the expected proportion in the general population. With BD diagnosis typically being made in early adulthood, this means that the excess burden of cognitive impairment will affect this population over several decades. There is a great need for effective interventions for cognitive dysfunction in BD, with significant potential to reduce adverse impacts on educational, occupational and domestic functioning over many years.

The wide variation in the prevalence estimates reported here calls attention to the need for greater consistency across studies. This could be achieved by using internationally recommended assessment batteries, such as those based on the MATRICS Consensus Cognitive Battery (Van Rheenen and Rossell, 2014; Yatham et al., 2010). Researchers should consider reporting impairment prevalence at more than one threshold, to facilitate comparison across studies. There is no single consensus threshold to define impairment in clinical practice, since this depends on contextual factors such as premorbid ability, but providing results for several relevant levels would maximise

value from data gathered. It would also be very useful to provide graphical summaries of score distributions, to indicate whether cognitive performance in BD samples (both BD-I and BD-II) demonstrates an overall distribution shift compared to healthy comparison groups, or a bimodal picture of distinct impaired versus preserved subgroups. Inspection of standardised mean differences together with proportions impaired does not permit a full appreciation of these issues. This review did not attempt to quantify the proportion of people with BD who are above average on cognitive measures, but this is arguably of equal importance in considering the diverse cognitive phenotype associated with this disorder. Finally, greater efforts should be made to recruit representative samples of adults with BD for cognitive studies. Much of the research in this field is carried out with clinic samples recruited by convenience. Only large, representative samples can provide an accurate picture of the burden of cognitive dysfunction in adults living with BD, through which we can understand the factors that influence risk and resilience.

REFERENCES

Altshuler, L.L., Ventura, J., van Gorp, W.G., Green, M.F., Theberge, D.C., Mintz, J., 2004. Neurocognitive function in clinically stable men with bipolar I disorder or schizophrenia and normal control subjects. Biol. Psychiatry 56, 560-569.

Arslan, F.C., Tiryaki, A., Ozkorumak, E., 2014. A comparison of euthymic bipolar patients with unaffected first-degree relatives and healthy controls in terms of neuropsychological functions. International Journal of Psychiatry in Clinical Practice 18, 208-214.

Arts, B., Jabben, N., Krabbendam, L., van Os, J., 2008. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. Psychol. Med. 38, 771-785.

Barrera, A., Vazquez, G., Tannenhaus, L., Lolich, M., Herbst, L., 2013. Theory of mind and functionality in bipolar patients with symptomatic remission. Revista De Psiquiatria Y Salud Mental 6, 67-74.

Bourne, C., Aydemir, O., Balanza-Martinez, V., Bora, E., Brissos, S., Cavanagh, J.T.O., Clark, L., Cubukcuoglu, Z., Dias, V.V., Dittmann, S., Ferrier, I.N., Fleck, D.E., Frangou, S., Gallagher, P., Jones, L., Kieseppa, T., Martinez-Aran, A., Melle, I., Moore, P.B., Mur, M., Pfennig, A., Raust, A., Senturk, V., Simonsen, C., Smith, D.J., Bio, D.S., Soeiro-de-Souza, M.G., Stoddart, S.D.R., Sundet, K., Szoke, A., Thompson, J.M., Torrent, C., Zalla, T., Craddock, N., Andreassen, O.A., Leboyer, M., Vieta, E., Bauer, M., Worhunsky, P.D., Tzagarakis, C., Rogers, R.D., Geddes, J.R., Goodwin, G.M., 2013. Neuropsychological testing of cognitive impairment in euthymic bipolar disorder: an individual patient data meta-analysis. Acta Psychiatr. Scand. 128, 149-162.

Cavanagh, J.T.O., Van Beck, M., Muir, W., Blackwood, D.H.R., 2002. Case-control study of neurocognitive function in euthymic patients with bipolar disorder: an association with mania. Br. J. Psychiatry 180, 320-326.

Cheung, E.Y.W., Halari, R., Cheng, K.M., Leung, S.K., Young, A.H., 2013. Cognitive performance is impaired in euthymic Chinese patients with Bipolar 1 Disorder. J. Affect. Disord. 151, 156-163.

Daban, C., Mathieu, F., Raust, A., Cochet, B., Scott, J., Etain, B., Leboyer, M., Bellivier, F., 2012. Is processing speed a valid cognitive endophenotype for bipolar disorder? J. Affect. Disord. 139, 98-101.

Doganavsargil-Baysal, O., Cinemre, B., Aksoy, U.M., Akbas, H., Metin, O., Fettahoglu, C., Gokmen, Z., Davran, F., 2013. Levels of TNF-, soluble TNF receptors (sTNFR1, sTNFR2), and cognition in bipolar disorder. Human Psychopharmacology-Clinical and Experimental 28, 160-167.

Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. British Medical Journal 315, 629-634.

Elshahawi, H.H., Essawi, H., Rabie, M.A., Mansour, M., Beshry, Z.A., Mansour, A.N., 2011. Cognitive functions among euthymic bipolar I patients after a single manic episode versus recurrent episodes. J. Affect. Disord. 130, 180-191.

Fakhry, H., El Ghonemy, S.H., Salem, A., 2013. Cognitive functions and cognitive styles in young euthymic patients with bipolar I disorder. J. Affect. Disord. 151, 369-377.

Ferrier, I.N., Stanton, B.R., Kelly, T.P., Scott, J., 1999. Neuropsychological function in euthymic patients with bipolar disorder. Br. J. Psychiatry 175, 246-251.

Frangou, S., Donaldson, S., Hadjulis, M., Landau, S., Goldstein, L.H., 2005. The Maudsley Bipolar Disorder Project: Executive dysfunction in bipolar disorder I and its clinical correlates. Biol. Psychiatry 58, 859-864.

Goswami, U., Sharma, A., Varma, A., Gulrajani, C., Ferrier, I.N., Young, A.H., Gallagher, P., Thompson, J.M., Moore, P.B., 2009. The neurocognitive performance of drug-free and medicated euthymic bipolar patients do not differ. Acta Psychiatr. Scand. 120, 456-463.

Ibrahim, N., Rahman, A.H.A., Shah, S.A., 2009. Verbal memory test performance in patients with bipolar I disorder attending a psychiatric clinic of a university hospital in Kuala Lumpur, Malaysia. ASEAN Journal of Psychiatry 10, 157-168.

Iverson, G.L., Brooks, B.L., Langenecker, S.A., Young, A.H., 2011. Identifying a cognitive impairment subgroup in adults with mood disorders. J. Affect. Disord. 132, 360-367.

Jamrozinski, K., Gruber, O., Kemmer, C., Falkai, P., Scherk, H., 2009. Neurocognitive functions in euthymic bipolar patients. Acta Psychiatr. Scand. 119, 365-374.

Juselius, S., Kieseppa, T., Kaprio, J., Lonnqvist, J., Tuulio-Henriksson, A., 2009. Executive Functioning in Twins with Bipolar I Disorder and Healthy Co-Twins. Archives of Clinical Neuropsychology 24, 599-606.

Kieseppa, T., Tuulio-Henriksson, A., Haukka, J., Van Erp, T., Glahn, D., Cannon, T.D., Partonen, T., Kaprio, J., Lonnqvist, J., 2005. Memory and verbal learning functions in twins with bipolar-I disorder, and the role of information-processing speed. Psychol. Med. 35, 205-215.

Lopera-Vasquez, J., Bell, V., López-Jaramillo, C., 2011. What is the contribution of executive dysfunction to the cognitive profile of bipolar disorder? A well-controlled direct comparison study. Revista Colombiana de Psiquiatría 40, 64S-75S.

Lopez-Jaramillo, C., Lopera-Vasquez, J., Gallo, A., Ospina-Duque, J., Bell, V., Torrent, C., Martinez-Aran, A., Vieta, E., 2010. Effects of recurrence on the cognitive performance of patients with bipolar I disorder: implications for relapse prevention and treatment adherence. Bipolar Disord. 12, 557-567.

Mann-Wrobel, M.C., Carreno, J.T., Dickinson, D., 2011. Meta-analysis of neuropsychological functioning in euthymic bipolar disorder: an update and investigation of moderator variables. Bipolar Disord. 13, 334-342.

Martino, D.J., Igoa, A., Marengo, E., Scapola, M., Strejilevich, S.A., 2011a. Neurocognitive Impairments and Their Relationship With Psychosocial Functioning in Euthymic Bipolar II Disorder. Journal of Nervous and Mental Disease 199, 459-464.

Martino, D.J., Strejilevich, S.A., Fassi, G., Marengo, E., Igoa, A., 2011b. Theory of mind and facial emotion recognition in euthymic bipolar I and bipolar II disorders. Psychiatry Research 189, 379-384.

Martino, D.J., Strejilevich, S.A., Marengo, E., Ibanez, A., Scapola, M., Igoa, A., 2014. Toward the identification of neurocognitive subtypes in euthymic patients with bipolar disorder. J. Affect. Disord. 167, 118-124.

Martino, D.J., Strejilevich, S.A., Scapola, M., Igoa, A., Marengo, E., Ais, E.D., Perinot, L., 2008. Heterogeneity in cognitive functioning among patients with bipolar disorder. J. Affect. Disord. 109, 149-156.

Martino, D.J., Strejilevich, S.A., Torralva, T., Manes, F., 2011c. Decision making in euthymic bipolar I and bipolar II disorders. Psychol. Med. 41, 1319-1327.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. J Clin Epidemiol 62, 1006-1012.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Systematic reviews 4, 1.

Munn, Z., Moola, S., Riitano, D., Lisy, K., 2014. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. International Journal of Health Policy Management 3, 123-128.

Mur, M., Portella, M.J., Martinez-Aran, A., Pifarre, J., Vieta, E., 2007. Persistent neuropsychological deficit in euthymic bipolar patients: Executive function as a core deficit. J Clin Psychiat 68, 1078-1086.

Normala, I., Abdul, H.A.R., Azlin, B., Nik Ruzyanei, N.J., Hazli, Z., Shah, S.A., 2010. Executive function and attention span in euthymic patients with bipolar 1 disorder. The Medical Journal of Malaysia 65, 199-203.

Osher, Y., Dobron, A., Belmaker, R.H., Bersudsky, Y., Dwolatzky, T., 2011. Computerized Testing of Neurocognitive Function in Euthymic Bipolar Patients Compared to Those with Mild Cognitive Impairment and Cognitively Healthy Controls. Psychotherapy and Psychosomatics 80, 298-303.

Pirkola, T., Tuulio-Henriksson, A., Glahn, D., Kieseppa, T., Haukka, J., Kaprio, J., Lonnqvist, J., Cannon, T.D., 2005. Spatial working memory function in twins with schizophrenia and bipolar disorder. Biol. Psychiatry 58, 930-936.

Robinson, L.J., Thompson, J.M., Gallagher, P., Goswami, U., Young, A.H., Ferrier, I.N., Moore, P.B., 2006. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. J. Affect. Disord. 93, 105-115.

Sánchez-Morla, E.M., Barabash, A., Martinez-Vizcaino, V., Tabares-Seisdedos, R., Balanza-Martinez, V., Cabranes-Diaz, J.A., Baca-Baldomero, E., Santos Gomez, J.L., 2009. Comparative study of neurocognitive function in euthymic bipolar patients and stabilized schizophrenic patients. Psychiatry Research 169, 220-228.

Sparding, T., Silander, K., Palsson, E., Ostlind, J., Sellgren, C., Ekman, C.J., Joas, E., Hansen, S., Landen, M., 2015. Cognitive Functioning in Clinically Stable Patients with Bipolar Disorder I and II. PLoS One 10, e0115562.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology - A proposal for reporting. Jama-J Am Med Assoc 283, 2008-2012.

Szmulewicz, A.G., Samame, C., Martino, D.J., Strejilevich, S.A., 2015. An updated review on the neuropsychological profile of subjects with bipolar disorder. Arch Clin Psychiat (São Paulo) 42, 139-146.

van der Werf-Eldering, M.J., Burger, H., Holthausen, E.A.E., Aleman, A., Nolen, W.A., 2010. Cognitive Functioning in Patients with Bipolar Disorder: Association with Depressive Symptoms and Alcohol Use. PLoS One 5, e13032.

Van Rheenen, T.E., Rossell, S.L., 2014. An empirical evaluation of the MATRICS Consensus Cognitive Battery in bipolar disorder. Bipolar Disord. 16, 318-325.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gotzsche, P.C., Vandenbroucke, J.P., Initiative, S., 2007. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiology 18, 800-804.

Yatham, L.N., Torres, I.J., Malhi, G.S., Frangou, S., Glahn, D.C., Bearden, C.E., Burdick, K.E., Martínez-Arán, A., Dittmann, S., Goldberg, J.F., Ozerdem, A., Aydemir, O., Chengappa, K.N.R., 2010. The International Society for Bipolar Disorders Battery for Assessment of Neurocognition (ISBD-BANC). Bipolar Disord. 12, 351-363.

FIGURE TITLES AND LEGENDS

Figure 1 PRISMA flow diagram showing results of literature search and screening.

Figure 2 Executive function impairment prevalence across different thresholds.

BADS, Behavioural Assessment of the Dysexecutive Syndrome; BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition; WCST, Wisconsin Card Sorting Test.

Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: sqrt((prevalence*(100-prevalence))/n). Results are stratified by impairment threshold (percentile), in descending order from least to most strict. Panel A shows executive measures whose scores do not have a prominent timed/speed contribution, and panel B shows executive measures whose scores are influenced by speed.

Figure 3 Attention/working memory impairment prevalence across different thresholds.

BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CPT, Continuous Performance Test; WAIS-III, Wechsler Adult Intelligence Scale third edition.

Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: sqrt((prevalence*(100-prevalence))/n). Results are stratified by impairment threshold (percentile), in descending order from least to most strict.

Figure 4 Speed/reaction time impairment prevalence across different thresholds.

BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CPT, Continuous Performance Test; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition.

Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: sqrt((prevalence*(100-prevalence))/n). Results are stratified by impairment threshold (percentile), in descending order from least to most strict.

Figure 5 Verbal and visual memory impairment prevalence across different thresholds.

BD, bipolar disorder; BD-I, bipolar disorder type I; CI, confidence interval; CNS-VS, Central Nervous System Vital Signs computerised battery; CVLT, California Verbal Learning Test; RCFT, Rey Complex Figure Test; WAIS-III, Wechsler Adult Intelligence Scale third edition.

Results include mixed BD and BD-I samples. Some studies reported results for several cognitive scores, and so there is sample overlap across rows. CI estimates are based on standard errors calculated as follows: sqrt((prevalence*(100-prevalence))/n). Results are stratified by impairment threshold (percentile), in descending order from least to most strict. Panel A shows verbal memory measures and panel B shows visual memory measures.

Table 1 Characteristics of included articles

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
Altshuler 2004	USA	40	22	BD-I DSM-III-R	HRSD <6 and YMRS <7 for 3 consecutive months	Head injury with LOC >1 hour; learning disability; migraine; liver function abnormalities; alcoholic dementia; abuse of alcohol in past 6 months; history of cocaine abuse/dependence; diabetes; hypertension; seizure disorder; any other neurologic illness; left-handedness; ECT in past 2 years; other current DSM-III-R Axis I disorder
Arslan 2014	Turkey	30	32	BD-I DSM-IV	HRSD <7 and MADRS <12 and YMRS ≤12	DSM Axis I comorbidities; mental retardation; hearing/visual loss interfering with clinical interview; alcohol/substance abuse in past 6 months; any disease affecting CNS; head trauma with LOC
Barrera 2013	Argentina	12	12	BD-I or BD-II Not stated	HRSD (17 items) <7 and YMRS <8	Significant medical diseases; neurological disorders; mental deficiency; drug abuse
Cavanagh 2002	UK	20	20	BD-I DSM-IV	HRSD ≤7 and MMS ≤2	Significant physical or neurological illness; stroke or head trauma; significant alcohol and/or drug misuse; ECT in past 6 months; comorbid psychiatric disorder; neurodegenerative disorder; learning disability; endocrine abnormalities
Cheung 2013	China (Hong Kong)	52	52	BD-I ICD-10 and DSM-IV	HRSD (21 items) <7 and YMRS <7 on two occasions 4 weeks apart	Mental retardation; change in psychotropic medication in past 4 weeks; DSM-IV alcohol/substance abuse in past 12 months; head injury with LOC; neurological disorder; history of psychiatric illness other than BD-I; significant physical health problem which could affect cognition
Daban 2012	France	53	60	BD DSM-III-R	MADRS <6 and BR-MRS <5	History of severe head trauma; learning difficulties; neurological disorder; current alcohol/drug abuse
Doganavsargil-Baysal 2013	Turkey	54	18	BD-I DSM-IV-TR	HRSD ≤7 and YMRS ≤5	Comorbid psychiatric or neurological disorders; IQ score <80; infectious or autoimmune diseases; on anti-inflammatory or antibiotic medication; biochemical values not within normal range
Elshahawi 2011	Egypt	50	50	BD-I or BD-II; history of ≥3 affective episodes ICD-10	HRSD <8 and YMRS <6	Comorbid psychiatric disorder; ECT in past 3 months; neurological disorder; mental retardation; substance abuse; organic cause of cognitive impairment
Fakhry 2013	UAE	30 (recent manic episode) 30 (recent depressive episode)	30	BD-I; history of ≤3 affective episodes; illness duration <5 years DSM-IV	MES and MAS <6; free from symptoms for at least 8 weeks and not fulfilling DSM-IV criteria for an affective episode	Comorbid psychiatric disorders; ECT in past 6 months; lithium-receiving patients in a trial
Ferrier 1999	UK	21 ('good' outcome) 20 ('poor' outcome)	20	BD-I; at least 5 years illness duration DSM-IV	HRSD ≤8 and MSS <20	Dementing disorder; learning disability; history of substance misuse, cerebrovascular disease, neurodegenerative disorders, head injury with concussion, clinical epilepsy, systemic illness with known cerebral consequences, severe hypertension, severe hepatic or renal disorder, or endocrine disorders other than corrected hypothyroidism

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
Frangou 2005	UK	44	44	BD-I DSM-IV	Syndromal remission: not meeting DSM-IV criteria for a mood episode for at least 3 months; no change in medication type/dose over the same period. Symptomatic remission: HRSD and MRS-SADS <10	None
Goswami 2009	India	22 (on medication) 22 (not on medication)	NA	BD-I DSM-IV	HRSD <8 and MSS <20 on two occasions 4 weeks apart	Other DSM Axis I or II diagnoses; cardiorespiratory, gastrointestinal, neurological and endocrine disorders (other than corrected hypothyroidism); substance misuse/dependency disorders; other medications e.g. anticholinergics, hypnotics or steroids
Ibrahim[a] 2009	Malaysia	40	40	BD-I DSM-IV	No active manic or depressive symptoms as reflected by YMRS and HRSD scores	Overtly disturbed/aggressive; severe mental retardation; dementia; significant CNS disease; head injury; comorbid psychiatric disorders; substance abuse/dependence; use of anticholinergics or benzodiazepines
Jamrozinski 2009	Germany	40	40	BD-I DSM-IV	MADRS ≤10 and YMRS ≤12	Other medical disorders
Juselius[b] 2009	Finland	26	114	BD-I DSM-IV (past diagnosis using ICD-8 or DSM-III-R)	In remission according to DSM-IV criteria	Other psychotic disorders; neurological disorders; brain injury; current alcohol dependence
Kieseppä[b] 2005	Finland	26	114	BD-I DSM-IV (past diagnosis using ICD-8 or DSM-III-R)	In full symptom remission according to DSM-IV criteria	Other psychotic disorders; neurological disorders; brain injury; current alcohol dependence
Lopera-Vásquez 2011	Colombia	40 (on medication) 31 (not on medication)	28	BD-I DSM-IV	ZSDS <8 and YMRS <6	Illicit substances or benzodiazepines in past 4 weeks; other psychiatric or neurological disorders; mental retardation; any treatment with ECT
López-Jaramillo 2010	Colombia	24 (1 manic episode) 27 (2 manic episodes) 47 (≥3 manic episodes)	66	BD-I DSM-IV	HRSD <8 and YMRS <6	Illicit substances or benzodiazepines in past 4 weeks; other psychiatric or neurological disorders that could affect cognition; mental retardation; any treatment with ECT; physical/sensory limitations that could affect performance
Martino[c] 2008	Argentina	50	30	BD-I or BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 6 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011a	Argentina	48 (BD-I) 39 (BD-II)	39	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011b	Argentina	45 (BD-I) 36 (BD-II)	34	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Martino[c] 2011c	Argentina	48 (BD-I) 37 (BD-II)	34	BD-I; BD-II DSM-IV	HRSD ≤8 and YMRS ≤6 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance

Author Year	Country	Sample *n*		BD sample type BD definition	Euthymia definition	Exclusion criteria
		BD	HC			
Martino[c] 2014	Argentina	100	40	BD-I or BD-II DSM-IV	HRSD ≤9 and YMRS ≤8 for at least 8 weeks	Substance abuse; mental retardation; neurological disease; any clinical condition that could affect cognitive performance
Mur 2007	Spain	44	46	BD-I or BD-II DSM-IV	HRSD (17-item) <8 and YMRS <6 for at least 3 months; on same treatment regimen and clinically stable for 3 months	Significant physical or neurologic illness; substance abuse/dependence in the past year; ECT in the past year; any mood-stabilising medication other than lithium
Normala[a] 2010	Malaysia	40	40	BD-I DSM-IV	No active manic or depressive symptoms as reflected by YMRS and HRSD scores	Overtly disturbed/aggressive; severe mental retardation; dementia; significant CNS disease; head injury; comorbid psychiatric disorders; substance abuse/dependence; use of anticholinergics or benzodiazepines
Osher 2011	Israel	51	495	BD-I DSM-IV	Consensus judgement by two clinicians based on full history and evidence of stability for at least three months[d]	Serious physical illness or substance abuse
Pirkola[b] 2005	Finland	22	100	BD-I DSM-III-R or DSM-IV	Not stated	Schizoaffective disorder; psychotic disorder other than BD-I; neurological disease; clinically significant head injury; mental retardation
Sánchez-Morla 2009	Spain	73	67	BD DSM-IV	HRSD <7 and YMRS <6 for 3 consecutive monthly evaluations	Neurological or medical diseases that can affect cognition; mental retardation; history of alcohol or other substance abuse/dependence in past 2 years; ECT in past 2 years; history of head injury with LOC
Sparding 2015	Sweden	64 (BD-I) 44 (BD-II)	86	BD-I; BD-II DSM-IV	MADRS and YMRS <14	None stated
van der Werf-Eldering 2010	The Netherlands	46	75	BD-I, BD-II or BD-NOS DSM-IV	IDS-SR <14 and YMRS <8	Mental retardation; systemic or neurological disease which could affect cognition; alcohol use disorder currently needing treatment in a specialised setting

BD, bipolar disorder; BD-I, bipolar disorder type I; BD-II, bipolar disorder type II; BD-NOS, bipolar disorder not otherwise specified; BR-MRS, Bech–Rafaelsen Mania Rating Scale; CNS, central nervous system; DSM, Diagnostic and Statistical Manual of Mental Disorders; DSM-III-R, Diagnostic and Statistical Manual of Mental Disorders third edition revised; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders fourth edition; DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders fourth edition text revision; ECT, electroconvulsive therapy; HC, healthy comparison; HRSD, Hamilton Rating Scale for Depression; ICD-8, International Classification of Diseases eighth revision; ICD-10, International Classification of Diseases tenth revision; IDS-SR, Inventory of Depressive Symptomatology - Self Rating; IQ, intelligence quotient; LOC, loss of consciousness; MADRS, Montgomery–Åsberg Depression Rating Scale; MAS, Bech–Rafaelsen Mania Scale; MES, Bech–Rafaelsen Melancholia Scale; MRS-SADS, Mania Rating Scale from the Schedule for Affective Disorders and Schizophrenia (Change Version); MSS, Bech's modification of Beigel's Mania State Rating Scale; NA, not applicable; YMRS, Young Mania Rating Scale; ZSDS, Zung Self-Rated Depression Scale.

[a] Studies contain overlapping samples.

[b] Studies contain overlapping samples.

[c] Studies contain overlapping samples.

[d] Information provided by author.

Table 2 Prevalence of cognitive impairment in BD-I samples[a]

Author Year	Sample *n*		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Altshuler 2004	40	22	WCST categories (executive)	Score 0 to 3	(42%)	(0%)	-0.98
			CVLT total recall 1-5 (verbal memory)	1.75 SD from published normative mean[b]	(22%)	(0%)	-0.99
Cavanagh 2002[c]	20	20	Stroop Color-Word Test (executive)	1.64 SD from HC mean	7 (36.8%)	(5%)[d]	-0.61
			Letter fluency (executive)		2 (10%)	(5%)[d]	-0.31
			BADS Six Elements (executive)		1 (5.3%)	(5%)[d]	-0.31
			CVLT trial 1 (verbal memory)		7 (35%)	(5%)[d]	-1.24
			CVLT total recall 1-5 (verbal memory)		5 (25%)	(5%)[d]	-1.06
			CVLT delayed recall (verbal memory)		8 (42.1%)	(5%)[d]	-0.96
			CVLT delayed recognition total (verbal memory)		4 (21.1%)	(5%)[d]	-0.62
			CVLT delayed recognition minus false positives (verbal memory)		4 (21.1%)	(5%)[d]	-0.66
			Stroop Color-Word Test (executive)	2 SD from HC mean	7 (36.8%)	(2.275%)[d]	-0.61
			Letter fluency (executive)		2 (10%)	(2.275%)[d]	-0.31
			BADS Six Elements (executive)		1 (5.3%)	(2.275%)[d]	-0.31
			CVLT trial 1 (verbal memory)		2 (10%)	(2.275%)[d]	-1.24
			CVLT total recall 1-5 (verbal memory)		4 (20%)	(2.275%)[d]	-1.06
			CVLT delayed recall (verbal memory)		5 (26.3%)	(2.275%)[d]	-0.96
			CVLT delayed recognition total (verbal memory)		4 (21.1%)	(2.275%)[d]	-0.62
			CVLT delayed recognition minus false positives (verbal memory)		4 (21.1%)	(2.275%)[d]	-0.66
Cheung 2013	52	52	CNS-VS neurocognition (overall)	5th percentile of published norm	(46.2%)	(0.0%)	-1.64
			CNS-VS executive function		(53.8%)	(0.0%)	-1.69
			CNS-VS cognitive flexibility		(57.7%)	(0.0%)	-1.66
			CNS-VS complex attention		(51.9%)	(1.9%)	-1.36
			CNS-VS processing speed		(26.9%)	(0.0%)	-1.21
			CNS-VS psychomotor speed		(30.8%)	(1.9%)	-1.15
			CNS-VS reaction time		(44.2%)	(13.5%)	-0.90
			CNS-VS memory composite		(30.8%)	(5.8%)	-0.80
			CNS-VS verbal memory		(28.8%)	(5.8%)	-0.71
			CNS-VS visual memory		(11.5%)	(3.8%)	-0.65
			1 SD from published normative mean on ≥2 index scores		(61.5%)	(1.9%)	NA
			2 SD from published normative mean on ≥2 index scores		(40.4%)	(0.0%)	NA
Fakhry 2013 S1: recent manic episode	30	30	MMSE (global)	Score <25	23 (76.7%)	0 (0%)	-4.62
			MTS (global)	Score <27	18 (60%)	0 (0%)	-2.10
			CDT (executive/visuospatial)	Score <6	0 (0%)	0 (0%)	-3.31
Fakhry 2013 S2: recent depressive episode	30	30	MMSE (global)	Score <25	6 (20%)	0 (0%)	-2.74
			MTS (global)	Score <27	5 (16.7%)	0 (0%)	-0.84
			CDT (executive/visuospatial)	Score <6	0 (0%)	0 (0%)	-2.89

Author Year	Sample *n*		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Ibrahim 2009[e] & Normala 2010[e]	40	40	Category fluency (executive/language)	Score ≤30	3 (7.5%)	0 (0%)[c]	-1.01
			TMT part A (speed/attention)[c]	>40/45/50 seconds[f]	19 (47.5%)	11 (27.5%)	-0.52
			TMT part B (executive)[c]	>90/100/135 seconds[f]	25 (62.5%)	13 (32.5%)	-0.81
			Digit span forward (attention)[c]	Span <5	3 (7.5%)	1 (2.5%)	-0.97
			Digit span backward (working memory)[c]	Span <4	18 (15.0%)	5 (12.5%)	-1.10
			RAVLT trial 1 (verbal memory)	Score <7	31 (77.5%)	13 (32.5%)	NR
			RAVLT trial 5 (verbal memory)	Score <12	23 (57.5%)	1 (2.5%)	NR
			RAVLT trials 1 to 5 (verbal memory)	Score increment <5	16 (40%)	3 (7.5%)	NR
			RAVLT list B (verbal memory)	Score <7	37 (92.5%)	14 (35%)	NR
Juselius 2009	26[g]	114	WCST categories (executive)	1.5 SD from HC mean	12 (50%)	(6.68%)[d]	-0.78
			WCST perseverative (executive)		13 (54%)	(6.68%)[d]	-1.74
			Stroop interference (executive)		15 (68%)	(6.68%)[d]	-3.58
			TMT B minus A (executive)		10 (42%)	(6.68%)[d]	-0.33
			Letter fluency (executive/language)		15 (63%)	(6.68%)[d]	-1.75
			Category fluency (executive/language)		18 (78%)	(6.68%)[d]	-3.40
Osher 2011[c]	51	495	Mindstreams global cognition	1.5 SD from HC mean	25 (49.0%)	(6.68%)[d]	-1.19
			Mindstreams executive function		13 (25.5%)	(6.68%)[d]	-0.83
			Mindstreams attention		20 (39.2%)	(6.68%)[d]	-1.04
			Mindstreams information processing speed		15 (29.4%)	(6.68%)[d]	-0.91
			Mindstreams memory		22 (43.1%)	(6.68%)[d]	-0.96
			Mindstreams verbal function		11 (21.6%)	(6.68%)[d]	-0.51
			Mindstreams visual-spatial		16 (31.4%)	(6.68%)[d]	-0.67
			Mindstreams motor skills		12 (23.5%)	(6.68%)[d]	-0.58

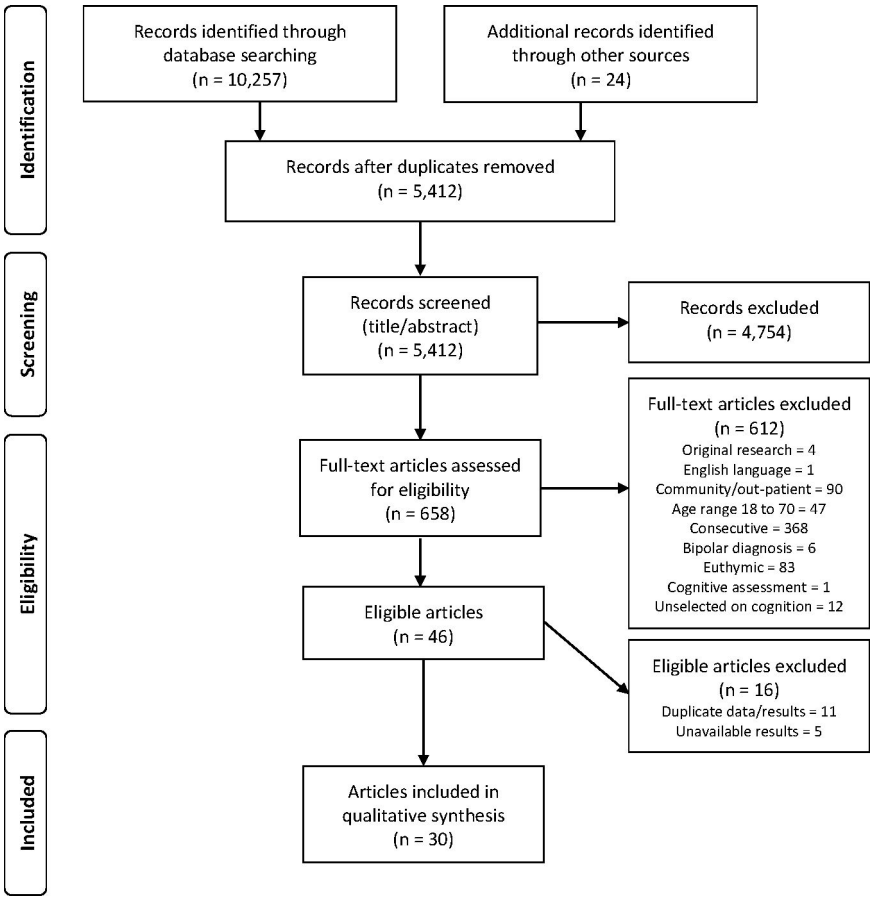
BADS, Behavioural Assessment of the Dysexecutive Syndrome; BD, bipolar disorder; BD-I, bipolar disorder type I; CDT, Clock Drawing Test; CNS-VS, Central Nervous System Vital Signs computerised battery; CVLT, California Verbal Learning Test; HC, healthy comparison; MMSE, Mini-mental State Examination; MTS, Mental Test Score; NA, not applicable; NR, unable to calculate as mean and SD not reported in article; RAVLT, Rey Auditory Verbal Learning Test; SD, standard deviation; TMT, Trailmaking Test; WCST, Wisconsin Card Sorting Test.
d is the standardised mean difference between BD and HC groups, calculated from unadjusted results in the article; negative values indicate worse performance in BD group.
[a] Sample characteristics are reported in Supplementary Table S1.
[b] T-score <32; impairment definition not explicit in article but inferred from bar graph of results.
[c] Prevalence data provided by author.
[d] By definition, according to impairment threshold applied.
[e] Same sample; RAVLT reported in Ibrahim 2009 and other tests reported in Normala 2010.
[f] Age groups 18-39, 40-49 and 50-59, respectively.
[g] Sample denominator for prevalence results ranges from 22 to 24.

Table 3 Prevalence of cognitive impairment in mixed BD samples[a]

Author Year	Sample *n*		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
Barrera 2013[b]	12	12	Reading the Mind in the Eyes test (theory of mind)	Score <21	6 (50%)	2 (16.7%)	-0.61
			Faux Pas Recognition Test cognitive items (theory of mind)	Score <0.75	7 (58.3%)	4 (33.3%)	-0.77
Daban 2012	53	60	WAIS-III Digit Symbol Coding (processing speed)	1.64 SD from HC mean	(30.2%)	(5%)[c]	-0.89
Martino 2014	100	40	Various tests (executive, attention/working memory, verbal memory, naming)	1.5 SD from published normative mean in ≥1 cognitive domain	(70%)	(27.5%)	NA
				2 SD from published normative mean in ≥2 cognitive domains	(30%)	(7.5%)	NA
Mur 2007[b]	44	46	TMT part B (executive)	1.5 SD from published normative mean	0 (0%)	0 (0%)	-0.72
			Letter fluency (executive/language)		6 (13.6%)	0 (0%)	-0.71
			WCST categories (executive)		18 (40.9%)	7 (15.2%)	-0.87
			WCST perseverative (executive)		15 (34.1%)	5 (10.9%)	-0.49
			Stroop inhibition (executive)		11 (25.0%)	1 (2.2%)	-1.30
			Digit span (attention/working memory)		3 (6.8%)	0 (0%)	NR
			TMT part A (speed/attention)		0 (0%)	0 (0%)	-0.28
			CVLT trial 1 (verbal memory)		11 (25.0%)	7 (15.2%)	0.19
			CVLT total words (verbal memory)		17 (38.6%)	6 (13.0%)	0.01
			CVLT immediate recall (verbal memory)		13 (29.5%)	4 (8.7%)	0.12
			CVLT delayed recall (verbal memory)		12 (27.3%)	6 (13.0%)	-0.33
			RCFT immediate (visual memory)		13 (29.5%)	0 (0%)	-0.52
			RCFT delayed (visual memory)		16 (36.4%)	4 (8.7%)	-0.55
Sánchez-Morla 2009	73	67	Executive composite z-score	1.64 SD from HC mean	33 (45.2%)	(5%)[c]	-1.80
			Sustained attention composite z-score		10 (13.7%)	(5%)[c]	-0.65
			Verbal memory composite z-score		21 (28.8%)	(5%)[c]	-1.18
			Visual memory composite z-score		24 (32.9%)	(5%)[c]	-1.10
			WCST % conceptual level response (executive)		(19.2%)	(5%)[c]	-1.02
			WCST % perseverative errors (executive)		(19.2%)	(5%)[c]	-1.01
			Stroop interference (executive)		(35.6%)	(5%)[c]	-0.98
			TMT part B (executive)		(32.9%)	(5%)[c]	-0.97
			Letter fluency (executive/language)		(16.4%)	(5%)[c]	-1.00
			Animal fluency (executive/language)		(24.7%)	(5%)[c]	-0.89
			Tower of Hanoi no. of movements (executive)		(19.0%)	(5%)[c]	-0.64
			Digit span backward (working memory)		(11.0%)	(5%)[c]	-0.53
			CPT hits (attention)		(9.6%)	(5%)[c]	-0.52
			CPT sensitivity A (attention)		(9.6%)	(5%)[c]	-0.58
			CPT reaction time (attention/speed)		(23.3%)	(5%)[c]	-0.72
			CVLT total recall 1-5 (verbal memory)		(19.2%)	(5%)[c]	-0.97
			CVLT short free-recall (verbal memory)		(27.4%)	(5%)[c]	-0.96
			CVLT long free-recall (verbal memory)		(15.1%)	(5%)[c]	-0.97
			CVLT short cued-recall (verbal memory)		(20.5%)	(5%)[c]	-1.11
			CVLT long cued-recall (verbal memory)		(23.3%)	(5%)[c]	-0.97

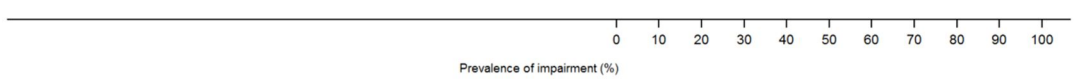
Author Year	Sample *n*		Cognitive measure	Impairment definition	Impairment prevalence *n* (%)		*d*
	BD	HC			BD	HC	
			CVLT recognition discriminability (verbal memory)		(8.2%)	(5%)[c]	-0.67
			CVLT semantic strategies trial A (verbal memory)		(41.1%)	(5%)[c]	-0.82
			RCFT copy (visuospatial)		(16.4%)	(5%)[c]	-0.51
			RCFT short-term (visual memory)		(31.5%)	(5%)[c]	-0.98
			RCFT long-term (visual memory)		(32.9%)	(5%)[c]	-1.01
van der Werf-Eldering 2010	46	75	Various tests (executive, attention/working memory, reaction time, verbal and visual memory)	2 SD from HC mean in ≥1 cognitive domain	6 (13%)	(2.275%)[c]	NA

BD, bipolar disorder; BD-I, bipolar disorder type I; CPT, Continuous Performance Test; CVLT, California Verbal Learning Test; HC, healthy comparison; NA, not applicable; NR, unable to calculate as mean and SD not reported in article; RCFT, Rey Complex Figure Test; SD, standard deviation; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition; WCST, Wisconsin Card Sorting Test.
d is the standardised mean difference between BD and HC groups, calculated from unadjusted results in the article; negative values indicate worse performance in BD group.
[a] Sample characteristics are reported in Supplementary Table S2.
[b] Prevalence data provided by author.
[c] By definition, according to impairment threshold applied.


Identification
Screening
Eligibility
Included
Records identified through database searching (n = 10,257)
Additional records identified through other sources (n = 24)
Records after duplicates removed (n = 5,412)
Records screened (title/abstract) (n = 5,412)
Records excluded (n = 4,754)
Full-text articles assessed for eligibility (n = 658)
Full-text articles excluded (n = 612)
Original research = 4
English language = 1
Community/out-patient = 90
Age range 18 to 70 = 47
Consecutive = 368
Bipolar diagnosis = 6
Euthymic = 83
Cognitive assessment = 1
Unselected on cognition = 12
Eligible articles (n = 46)
Eligible articles excluded (n = 16)
Duplicate data/results = 11
Unavailable results = 5
Articles included in qualitative synthesis (n = 30)

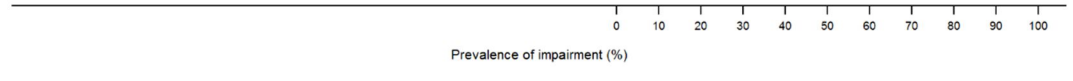
A

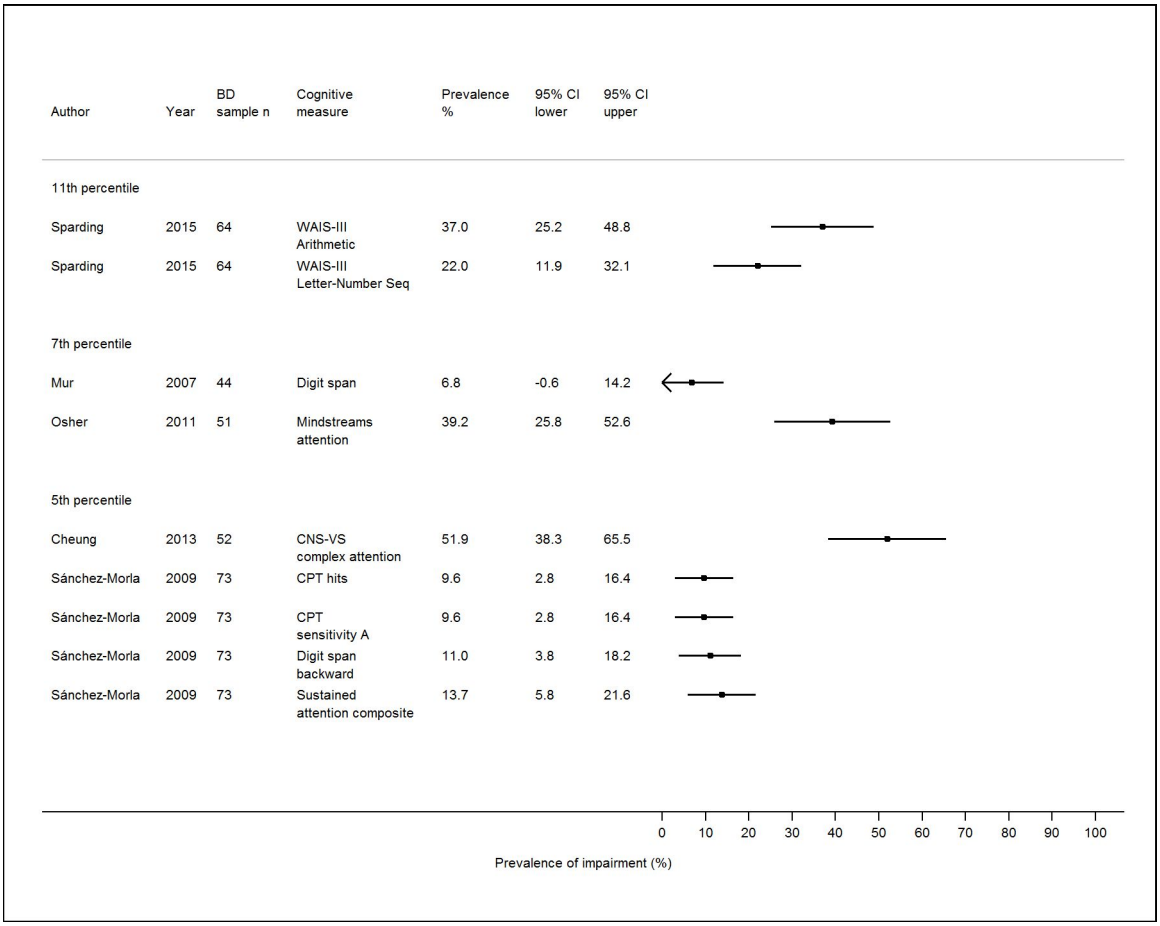
Author	Year	BD sample n	Cognitive measure	Prevalence %	95% CI lower	95% CI upper
11th percentile						
Sparding	2015	64	Tower rule violations	34.0	22.4	45.6
Sparding	2015	64	Verbal Fluency switch accuracy	18.0	8.6	27.4
Sparding	2015	64	WAIS-III Matrix Reasoning	21.0	11.0	31.0
Sparding	2015	64	WAIS-III Similarities	13.0	4.8	21.2
7th percentile						
Mur	2007	44	WCST categories	40.9	26.4	55.4
Juselius	2009	24	WCST categories	50.0	30.0	70.0
Mur	2007	44	WCST perseverative	34.1	20.1	48.1
Juselius	2009	24	WCST perseverative	54.0	34.1	73.9
5th percentile						
Cavanagh	2002	19	BADS Six Elements	5.3	-4.8	15.4
Cheung	2013	52	CNS-VS cognitive flexibility	57.7	44.3	71.1
Cheung	2013	52	CNS-VS executive function	53.8	40.2	67.4
Sánchez-Morla	2009	73	Executive composite	45.2	33.8	56.6
Sánchez-Morla	2009	73	Tower of Hanoi movements	19.0	10.0	28.0
Sánchez-Morla	2009	73	WCST conceptual	19.2	10.2	28.2
Sánchez-Morla	2009	73	WCST perseverative	19.2	10.2	28.2
2nd percentile						
Cavanagh	2002	19	BADS Six Elements	5.3	-4.8	15.4

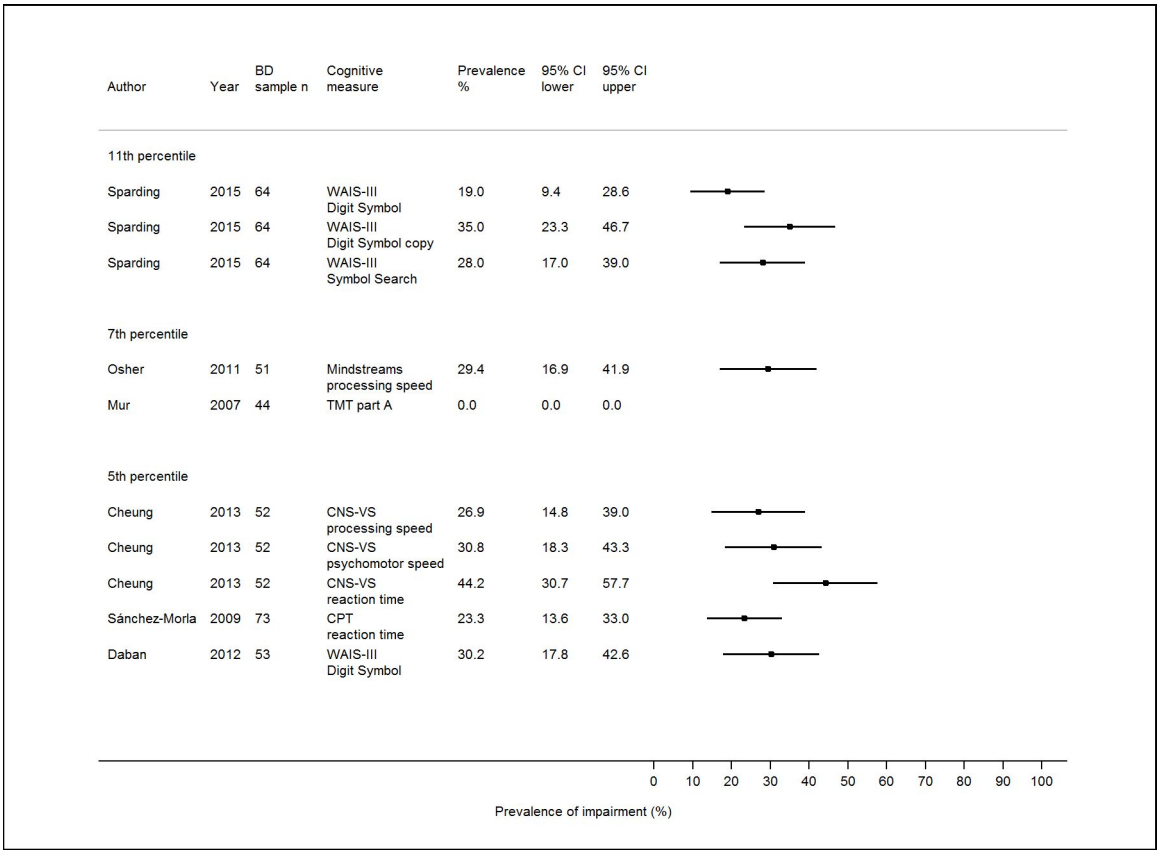


B

Author	Year	BD sample n	Cognitive measure	Prevalence %	95% CI lower	95% CI upper
11th percentile						
Sparding	2015	64	Design Fluency 3 switch	32.0	20.6	43.4
Sparding	2015	64	Stroop Color-Word 3 inhib	29.0	17.9	40.1
Sparding	2015	64	Stroop Color-Word 4 inh/sw	27.0	16.1	37.9
Sparding	2015	64	TMT 2 number seq	34.0	22.4	45.6
Sparding	2015	64	TMT 3 letter seq	22.0	11.9	32.1
Sparding	2015	64	TMT 4 switch	48.0	35.8	60.2
Sparding	2015	64	Verbal Fluency category	38.0	26.1	49.9
Sparding	2015	64	Verbal Fluency switch total	34.0	22.4	45.6
7th percentile						
Juselius	2009	23	Category fluency	78.0	61.1	94.9
Mur	2007	44	Letter fluency	13.6	3.5	23.7
Juselius	2009	24	Letter fluency	63.0	43.7	82.3
Osher	2011	51	Mindstreams executive function	25.5	13.5	37.5
Mur	2007	44	Stroop inhibition	25.0	12.2	37.8
Juselius	2009	22	Stroop Interference	68.0	48.5	87.5
Juselius	2009	24	TMT B minus A	42.0	22.3	61.7
Mur	2007	44	TMT part B	0.0	0.0	0.0
5th percentile						
Sánchez-Morla	2009	73	Category fluency	24.7	14.8	34.6
Cavanagh	2002	20	Letter fluency	10.0	-3.1	23.1
Sánchez-Morla	2009	73	Letter fluency	16.4	7.9	24.9
Cavanagh	2002	19	Stroop Color-Word	36.8	15.1	58.5
Sánchez-Morla	2009	73	Stroop Interference	35.6	24.6	46.6
Sánchez-Morla	2009	73	TMT part B	32.9	22.1	43.7
2nd percentile						
Cavanagh	2002	20	Letter fluency	10.0	-3.1	23.1
Cavanagh	2002	19	Stroop Color-Word	36.8	15.1	58.5

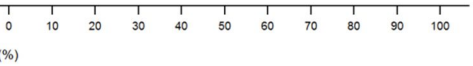



Author
Year
BD sample n
Cognitive measure
Prevalence %
95% CI lower
95% CI upper
11th percentile
Sparding 2015 64 WAIS-III Arithmetic 37.0 25.2 48.8
Sparding 2015 64 WAIS-III Letter-Number Seq 22.0 11.9 32.1
7th percentile
Mur 2007 44 Digit span 6.8 -0.6 14.2
Osher 2011 51 Mindstreams attention 39.2 25.8 52.6
5th percentile
Cheung 2013 52 CNS-VS complex attention 51.9 38.3 65.5
Sánchez-Morla 2009 73 CPT hits 9.6 2.8 16.4
Sánchez-Morla 2009 73 CPT sensitivity A 9.6 2.8 16.4
Sánchez-Morla 2009 73 Digit span backward 11.0 3.8 18.2
Sánchez-Morla 2009 73 Sustained attention composite 13.7 5.8 21.6
0 10 20 30 40 50 60 70 80 90 100
Prevalence of impairment (%)


Author
Year
BD sample n
Cognitive measure
Prevalence %
95% CI lower
95% CI upper
11th percentile
Sparding 2015 64 WAIS-III Digit Symbol 19.0 9.4 28.6
Sparding 2015 64 WAIS-III Digit Symbol copy 35.0 23.3 46.7
Sparding 2015 64 WAIS-III Symbol Search 28.0 17.0 39.0
7th percentile
Osher 2011 51 Mindstreams processing speed 29.4 16.9 41.9
Mur 2007 44 TMT part A 0.0 0.0 0.0
5th percentile
Cheung 2013 52 CNS-VS processing speed 26.9 14.8 39.0
Cheung 2013 52 CNS-VS psychomotor speed 30.8 18.3 43.3
Cheung 2013 52 CNS-VS reaction time 44.2 30.7 57.7
Sánchez-Morla 2009 73 CPT reaction time 23.3 13.6 33.0
Daban 2012 53 WAIS-III Digit Symbol 30.2 17.8 42.6
0 10 20 30 40 50 60 70 80 90 100
Prevalence of impairment (%)

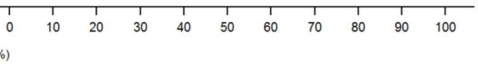
A

Author	Year	BD sample n	Cognitive measure	Prevalence %	95% CI lower	95% CI upper
7th percentile						
Mur	2007	44	CVLT long free recall	27.3	14.1	40.5
Mur	2007	44	CVLT short free recall	29.5	16.0	43.0
Mur	2007	44	CVLT total recall 1-5	38.6	24.2	53.0
Mur	2007	44	CVLT trial 1	25.0	12.2	37.8
5th percentile						
Cheung	2013	52	CNS-VS verbal memory	28.8	16.5	41.1
Cavanagh	2002	19	CVLT del recog - false pos	21.1	2.8	39.4
Cavanagh	2002	19	CVLT delay recog total	21.1	2.8	39.4
Sánchez-Morla	2009	73	CVLT long cued recall	23.3	13.6	33.0
Sánchez-Morla	2009	73	CVLT long free recall	15.1	6.9	23.3
Cavanagh	2002	19	CVLT long free recall	42.1	19.9	64.3
Sánchez-Morla	2009	73	CVLT recognition discrim	8.2	1.9	14.5
Sánchez-Morla	2009	73	CVLT semantic A	41.1	29.8	52.4
Sánchez-Morla	2009	73	CVLT short cued recall	20.5	11.2	29.8
Sánchez-Morla	2009	73	CVLT short free recall	27.4	17.2	37.6
Sánchez-Morla	2009	73	CVLT total recall 1-5	19.2	10.2	28.2
Cavanagh	2002	20	CVLT total recall 1-5	25.0	6.0	44.0
Cavanagh	2002	20	CVLT trial 1	35.0	14.1	55.9
Sánchez-Morla	2009	73	Verbal memory composite	28.8	18.4	39.2
4th percentile						
Altshuler	2004	40	CVLT total recall 1-5	22.0	9.2	34.8
2nd percentile						
Cavanagh	2002	19	CVLT del recog - false pos	21.1	2.8	39.4
Cavanagh	2002	19	CVLT delay recog total	21.1	2.8	39.4
Cavanagh	2002	19	CVLT long free recall	26.3	6.5	46.1
Cavanagh	2002	20	CVLT total recall 1-5	20.0	2.5	37.5
Cavanagh	2002	20	CVLT trial 1	10.0	-3.1	23.1



B

Author	Year	BD sample n	Cognitive measure	Prevalence %	95% CI lower	95% CI upper
11th percentile						
Sparding	2015	64	RCFT 3-minute recall	23.0	12.7	33.3
Sparding	2015	64	RCFT recognition	22.0	11.9	32.1
Sparding	2015	64	WAIS-III Digit Symbol pairing	26.0	15.3	36.7
Sparding	2015	64	WAIS-III Digit Symbol recall	25.0	14.4	35.6
7th percentile						
Mur	2007	44	RCFT delayed	36.4	22.2	50.6
Mur	2007	44	RCFT immediate	29.5	16.0	43.0
5th percentile						
Cheung	2013	52	CNS-VS visual memory	11.5	2.8	20.2
Sánchez-Morla	2009	73	RCFT long term	32.9	22.1	43.7
Sánchez-Morla	2009	73	RCFT short term	31.5	20.8	42.2
Sánchez-Morla	2009	73	Visual memory composite	32.9	22.1	43.7



Appendix 1 Search strategy implemented in Web of Science

Search set-up: English; publication year 1994 to date

1	TS=((bipolar NEAR/3 depress*) OR (bipolar NEAR/3 disorder*) OR (manic NEAR/0 depress*))
2	TS=(cogniti* OR neuro-cogniti* OR neurocogniti* OR neuro-psycholog* OR neuropsycholog* OR speed OR reaction OR attention OR memory OR learning OR *spatial OR executive OR reasoning OR IQ OR intelligence)
3	TS=(impair* OR dysfunction* OR declin* OR deteriorat* OR defici*)
4	(#1 AND #2 AND #3)
5	TI=(therapy OR CBT OR cognitive-behavio* OR (cognitive NEAR/0 behavio*))
6	#4 NOT #5
7	TS=((nursing NEAR/0 home*) OR (care NEAR/0 home*))
8	#6 NOT #7
9	TS=(dement*)
10	#8 NOT #9

Results limits: Refined by: Databases: (WOS OR SCIELO OR MEDLINE OR CCC OR DRCI) AND DOCUMENT TYPES: (ARTICLE OR CLINICAL TRIAL OR DATA SET OR UNSPECIFIED OR OTHER OR DATA STUDY) AND LANGUAGES: (ENGLISH)

Appendix 2 Eligibility checklist used when screening titles, abstracts and full text

No.	Criteria and definitions
1	**Original research published in peer-reviewed journals, from 1994 until date of search** · Not editorials, opinion papers, reviews, meta-analyses (but 'individual patient data meta/mega-analyses' that involve re-analysis of raw data [rather than group effect sizes] are acceptable) · Not conference proceedings, books, book chapters, academic theses · Not single case studies or case series (must be group studies) · If unsure whether journal is peer-reviewed, check the journal website or look up the title at https://ulrichsweb.serialssolutions.com/ and look for the 'refereed' symbol:  · Publication year is 1994 or later
2	**Articles published in English** · Full text of the article must be available in English
3	**Studies of community-dwelling adults aged 18 to 70 years inclusive** · Out-patient service or general population setting (this refers to the setting in which participants were recruited and/or the study took place) · BD participants may be recruited from clinic attendees, or from a population study (e.g. large-scale population registry) · Participants must have been living in the community at time of assessment · Not currently a hospital in-patient (but hospital/clinic out-patient is acceptable) · Not living in a nursing/care home · Participants are adults of minimum age 18 years and maximum age 70 years (abstracts describing samples as 'adolescent' or 'elderly' are presumed to be ineligible unless they also mention a separate sample of 'adults' in the study) · There may be a healthy comparison group in the study, but this is not a requirement
4	**Cross-sectional study design** · The key criterion is that the assessment of mood disorder status (whether the person has euthymic BD – the 'exposure') and the assessment of cognitive performance (the 'outcome') took place at the same time (maximum allowable time gap between establishing euthymic status and carrying out cognitive assessment = 2 weeks) · The mood disorder may have been diagnosed in the past; this is acceptable (as long as the confirmation of current euthymic status took place at the same time as the cognitive assessment – see section 8 below) · A cross-sectional study of this type may be part of a prospective/longitudinal/cohort study or a treatment trial; this is acceptable as long as separate (standalone) results are presented for the cross-sectional component that we are interested in · Studies investigating whether cognitive status predicts *future* remission of mood disorder are not eligible
5	**If clinical out-patient service setting, samples must be consecutively recruited** · Where BD participants have been recruited via a clinical service, recruitment must have been consecutive and representative of the target group · This means participants should not have been selectively approached (e.g. on the basis of their cognitive function or some other characteristic) in a way that might bias the results of the study · All eligible patients in the target group should have had an equal chance of being approached · For example, if the target group was patients with euthymic BD, and the researchers considered all such patients attending the clinic between time X and time Y (or a randomly chosen subset of these), then that would be acceptable
6	**Primary diagnosis of bipolar disorder (BD)** · Primary means the disorder has been diagnosed in its own right, not secondary to another illness · BD = History of bipolar disorder (type I or II or unspecified), meeting DSM or ICD criteria · Evidence for meeting diagnostic criteria may come from direct assessment as part of the study (often including a semi-structured interview schedule), or from a doctor's diagnosis recorded in the medical notes · Questionnaire measures of mood state alone (without reference to a diagnostic reference system) are not acceptable · The diagnosis may have been made at any time in the person's life up until the time of the study
7	**Euthymic at time of assessment** · Not meeting DSM or ICD criteria for a depressive or manic episode at time of cognitive

	assessment; or as otherwise defined by the study authors based on an appropriate clinical measure · For example, the authors may define euthymia as being below X threshold on a depression or mania rating scale · The concept of euthymia may be referred to as remitted/remission, or recovery, or stable on treatment, or treatment-responsive (or other similar phrase) · Non-euthymic patients may be described as acutely manic/unwell, or treatment-resistant (or other similar phrase) · Baseline samples within treatment trial studies are presumed to be NOT euthymic unless the abstract says otherwise
8	**Assessed using at least one standardised cognitive measure** · The measure should be an objective test, on which the participant's performance is assessed directly · Self-report questionnaires are not acceptable (e.g. the participant rates how good they think their cognition is) · Informant-rated questionnaires are not acceptable (e.g. the participant's spouse rates how good they think the participant's cognition is) · Informal behavioural observations are not acceptable (e.g. researcher observes participant without using any standardised rating scale) · The cognitive test may cover one or more of the following abilities/domains: o Global/overall function (cognitive/neuropsychological) o Processing speed/psychomotor speed/reaction time o Attention/vigilance/alertness/concentration o Working memory/memory/learning/encoding/recall/recognition/retrieval o Spatial/visuospatial ability o Language/naming/comprehension o Executive function (including planning, strategy-formation, problem-solving, decision-making, initiation, self-monitoring, self-regulation, mental control, goal management, goal neglect, inhibition, response suppression, fluency, word generation, perseveration, set-shifting, rule-shifting, flexibility, impulsivity, sequencing, dual-tasking, multi-tasking) o Reasoning/abstraction/concept formation/IQ/intelligence o Social cognition (including theory of mind, meta-cognition – e.g. judgement of other people's thoughts/behaviours) · Studies of non-conscious learning (e.g. classical conditioning and extinction) are not eligible · Studies using only experimental neuroimaging tasks (e.g. oddball/continuous performance tasks that are not interpretable in their own right) are not eligible · Studies of basic emotional processing only, without an explicit social cognition aspect, (e.g. reaction times to emotional faces) are not eligible · The test should yield a numeric score, or a rating (e.g. pass/fail, or poor/fair/good, or impaired/unimpaired) · *NOTE*: The mood disorder literature also contains many studies about cognitive distortions/biases. For example, a study may look at biased thinking patterns, attitudes, or rumination. This can be thought of as *cognitive style*, which is not the same as cognitive function in the neuropsychological sense outlined above. Therefore studies which are solely about these cognitive distortions, including treatments such as CBT aimed at changing these distortions, are not relevant to this review. · Similarly, some studies may focus on *formal thought disorder* (e.g. tangentiality, flight of ideas); these are not eligible unless a cognitive assessment of the type outlined above is also conducted.
9	**Samples NOT selected on the basis of presence of cognitive impairment (known or suspected)** · Participants (BD participants, or healthy comparison participants) should have been recruited (approached/selected) based on their exposure status (having or not having BD), and NOT based on their outcome status (having or not having cognitive impairment) · It is possible that potential participants who were approached to take part might have been more or less likely to agree depending on their cognitive status; this is outside the researcher's control and so is not the focus here. Rather, the issue is that the *initial approach/selection by the researcher* should not be based on cognitive status

Appendix 3 Instructions for data extraction

Spreadsheet field no.	Description	Entry format
1	Publication ID	Number
2	Sub-ID (use if study contains separate results for >1 BD sample)	Number-letter
3	Authors (e.g. Bloggs, A.B., Jones, C.D. & Smith, E.F.)	Free text
4	Year	XXXX
5	Journal	Free text
6	Volume	Free text
7	Pages	XX-YY
8	Title	Free text
9	Corresponding author's name and email	Free text
10	Study funding source	Free text or Not stated
11	Any other conflicts of interest declared	Free text or Not stated
12	Study setting: type	1=general community (e.g. population cohort) 2=clinical service 3=other 4=not stated
13	Study setting: details	Free text or Not stated
14	Population from which BD sample was drawn (e.g. all BD patients known to clinical service between date X and date Y; all families with 2 or more BD patients, etc)	Free text or Not stated
15	Healthy comparison group included	1=none (no healthy controls) 2=internal (from same population e.g. BD and controls taken from same general pop cohort) 3=external (from different population e.g. BD from clinic and controls from posters in community)
16	Comparison group matching (NB this refers to planned *matching during recruitment* – i.e. matched study *design* - not whether some characteristics happened to be similar after analysis)	1=not matched 2=matched to BD at group level only (unpaired design) 3=matched to BD individually (paired design) 4=NA 5=not stated
17	Comparison group: characteristics matched by	Free text or NA or Not stated
18	Recruitment procedure for BD group	Free text or Not stated
19	Recruitment procedure for comparison group	Free text or NA or Not stated
20	Inclusion criteria for BD group	Free text or Not stated
21	Exclusion criteria for BD group	Free text or Not stated
22	Inclusion criteria for comparison group	Free text or NA or Not stated
23	Exclusion criteria for comparison group	Free text or NA or Not stated
24	Informed consent obtained from all participants	1=yes 2=no 3=not stated
25	Power/sample size calculation reported for the study	1=yes 2=no
26	Country where study took place	Free text or Not stated
27	Language in which cognitive test was administered (assume English if test name and citation are for English version AND study is from English-speaking country)	Free text or Not stated
28	BD sample: sample size	n or Not stated
29	BD sample: age (mean and SD, or median and IQR, or range)	Free text or Not stated
30	BD sample: sex (n and % male)	Free text or Not stated
31	BD sample: ethnicity (n and % in each category)	Free text or Not stated

32	BD sample: education (info re years or qualifications)	Free text or Not stated
33	BD sample: socioeconomic status (how measured, and status of sample)	Free text or Not stated
34	Comparison sample: sample size	n or NA or Not stated
35	Comparison sample: age (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
36	Comparison sample: sex (n and % male)	Free text or NA or Not stated
37	Comparison sample: ethnicity (n and % in each category)	Free text or NA or Not stated
38	Comparison sample: education (info re years or qualifications)	Free text or NA or Not stated
39	Comparison sample: socioeconomic status (how measured, and status of sample)	Free text or NA or Not stated
40	Qualifications/training of person who made BD diagnosis	Free text or Not stated
41	Qualifications/training of person who did cognitive assessment	Free text or Not stated
42	BD definition (e.g. DSM IV/ICD-10 criteria)	Free text or Not stated
43	Euthymia definition (e.g. score less than X on named questionnaire, or clinician judgement of remission, etc)	Free text or Not stated
44	Euthymia confirmed at time of cognitive assessment	1=yes 2=no 3=not stated
45	Age at onset of BD (mean and SD, or median and IQR, or range)	Free text or Not stated
46	Duration since onset of BD (mean and SD, or median and IQR, or range)	Free text or Not stated
47	Number of previous illness episodes, all types combined (mean and SD, or median and IQR, or range)	Free text or Not stated
48	Number of previous depressive episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
49	Number of previous manic/hypomanic episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
50	Number of previous mixed episodes (mean and SD, or median and IQR, or range)	Free text or Not stated
51	BD sample: n and % currently taking any psychotropic medication	Free text or Not stated
52	BD sample: further details of medication type and dosage	Free text or Not stated
53	Comparison sample: n and % currently taking any psychotropic medication	Free text or NA or Not stated
54	Comparison sample: further details of medication type and dosage	Free text or NA or Not stated
55	Cognitive assessment: name of test	Free text
56	Cognitive assessment: source reference (journal citation; or name of publishing company)	Free text or Not stated
57	Cognitive assessment: domain covered (e.g. verbal memory)	Free text or Not stated
58	Cognitive assessment: definition of impairment threshold (e.g. 2SD below control mean score on one test; or 2SD below control mean score on at least two tests; or 1.5SD below published norms; or scores less than X; etc)	Free text or Not stated
59	Prevalence of cognitive impairment in BD sample (n and % who meet the stated definition of impairment given in field 58)	Free text or Not stated

60	Prevalence of cognitive impairment in comparison sample (n and % who meet the stated definition of impairment given in field 58)	Free text or NA or Not stated
61	Source of prevalence info (or explanation for missing info)	1=provided in paper 2=author provided on request 3=info missing: author replied but did not provide 4=info missing: author did not reply
62	Statistical analysis method to test/adjust for confounders/covariates (e.g. multiple regression)	Free text or NA or Not stated
63	Confounders/covariates (apart from other cognitive measures) that were significantly associated with cognitive performance in the analysis (e.g. higher age was a significant independent predictor of worse performance)	Free text or NA or Not stated
64	Missing data: brief description of level of missing data in the analysis (e.g. 10% of BD patients were missing a memory score)	Free text or Not stated
65	Missing data: brief description of how authors dealt with missing data (e.g. complete case analysis, or all available data, or imputation)	Free text or Not stated
66	Any other comments	Free text
67 *	Group cognitive score: BD sample (unadjusted group average) (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
68 *	Group cognitive score: comparison sample (unadjusted group average) (mean and SD, or median and IQR, or range)	Free text or NA or Not stated
69 *	Cohen's *d* (if M and SD available in fields 67 and 68). Minus sign = BD worse; calculate using http://www.uccs.edu/~lbecker/	Free text or NA or Not stated
70	Full list of all confounders/covariates tested (apart from other cognitive measures), whether significant or not (additional info for field 63)	Free text or NA or Not stated

* Optional fields – only complete these for papers for which information is available regarding prevalence of impairment (fields 59 and 60), otherwise enter NA

Appendix 4 Supplementary references for eligible articles not included in synthesis

1. Antila M, Partonen T, Kieseppa T, Suvisaari J, Eerola M, Lonnqvist J, et al. Cognitive functioning of bipolar I patients and relatives from families with or without schizophrenia or schizoaffective disorder. J Affect Disord. 2009;116(1-2):70-9.
2. Antila M, Tuulio-Henriksson A, Kieseppa T, Eerola M, Partonen T, Lonnqvist J. Cognitive functioning in patients with familial bipolar I disorder and their unaffected relatives. Psychol Med. 2007;37(5):679-87.
3. Braw Y, Sitman R, Sela T, Erez G, Bloch Y, Levkovitz Y. Comparison of insight among schizophrenia and bipolar disorder patients in remission of affective and positive symptoms: Analysis and critique. European Psychiatry. 2012;27(8):612-8.
4. Donaldson S, Goldstein LH, Landau S, Raymont V, Frangou S. The Maudsley Bipolar Disorder Project: The effect of medication, family history, and duration of illness on IQ and memory in bipolar I disorder. J Clin Psychiat. 2003;64(1):86-93.
5. Frangou S. The Maudsley Bipolar Disorder Project. Epilepsia. 2005;46:19-25.
6. Goswami U, Sharma A, Khastigir U, Ferrier IN, Young AH, Gallagher P, et al. Neuropsychological dysfunction, soft neurological signs and social disability in euthymic patients with bipolar disorder. Br J Psychiatry. 2006;188:366-73.
7. Haldane M, Frangou S. The Maudsley Bipolar Disorder Project: Insights into the role of the prefrontal cortex in bipolar disorder I. Revista de Psiquiatria do Rio Grande do Sul. 2005;27(3):1-20.
8. Kauer-Sant'Anna M, Yatham LN, Tramontina J, Weyne F, Cereser KM, Gazalle FK, et al. Emotional memory in bipolar disorder. Br J Psychiatry. 2008;192(6):458-63.
9. Martinez-Aran A, Penades R, Vieta E, Colom F, Reinares M, Benabarre A, et al. Executive function in patients with remitted bipolar disorder and schizophrenia and its relationship with functional outcome. Psychotherapy and Psychosomatics. 2002;71(1):39-46.
10. Mur M, Portella MJ, Martinez-Aran A, Pifarre J, Vieta E. Influence of Clinical and Neuropsychological Variables on the Psychosocial and Occupational Outcome of Remitted Bipolar Patients. Psychopathology. 2009;42(3):148-56.
11. Palsson E, Figueras C, Johansson AGM, Ekman C-J, Hultman B, Ostlind J, et al. Neurocognitive function in bipolar disorder: a comparison between bipolar I and II disorder and matched controls. BMC Psychiatry. 2013;13:165.
12. Strejilevich SA, Martino DJ. Cognitive function in adulthood and elderly euthymic bipolar patients: A comparison to test models of cognitive evolution. J Affect Disord. 2013;150(3):1188-91.
13. Torralva T, Gleichgerrcht E, Torrente F, Roca M, Strejilevich SA, Cetkovich M, et al. Neuropsychological functioning in adult bipolar disorder and ADHD patients: A comparative study. Psychiatry Research. 2011;186(2-3):261-6.
14. Trivedi JK, Goel D, Sharma S, Singh AP, Sinha PK, Tandon R. Cognitive functions in stable schizophrenia & euthymic state of bipolar disorder. Indian Journal of Medical Research. 2007;126(5):433-9.
15. van Gorp WG, Altshuler L, Theberge DC, Mintz J. Declarative and procedural memory in bipolar disorder. Biol Psychiatry. 1999;46(4):525-31.
16. van Gorp WG, Altshuler L, Theberge DC, Wilkins J, Dixon W. Cognitive impairment in euthymic bipolar patients with and without prior alcohol dependence - A preliminary study. Arch Gen Psychiatry. 1998;55(1):41-6.

Table S1 Characteristics of BD-I samples

Author Year	Age (years), M (SD)		Male gender, *n* (%)		Education (years), M (SD)		BD onset age and/or duration (years), M (SD)	BD past illness episodes, M (SD)	Psychotropic medication in BD sample
	BD	HC	BD	HC	BD	HC			
Altshuler 2004	49.9 (13.9)	51.8 (12.6)	40 (100%)	22 (100%)	15.5 (2.4)	14.9 (2.0)	Onset 26.7 (9.2) Duration 24.9 (11.1)	Not stated	Lithium 63%, anticonvulsant 30%, antidepressant 10%, benzodiazepine 8%, typical antipsychotic 15%, anticholinergic 5%
Cavanagh 2002	43.6 (14.2)	42.2 (14.7)	10 (50%)	10 (50%)	Not stated	Not stated	Duration 16 (12.5)	Depressive 6 (6) Manic 6 (7)	SSRI 35%, typical antipsychotic 40%, lithium 40%, carbamazepine 25%
Cheung 2013	38.57 (10.70)	37.76 (10.27)	19 (36.5%)	19 (36.5%)	12.0 (2.94)	14.04 (3.11)	Onset 24.63 (7.6) Duration 13.3 (8.3)	Depressive 5.1 (5.2) Manic 5.2 (5.0)	Monotherapy (sodium valproate, lithium, antipsychotic, carbamazepine or lamotrigine) 48.1%, combination therapy 48.1% (inc. anticholinergic 15.3%); Two patients received short-acting low-dose benzodiazepine 12h before assessment
Fakhry 2013 S1: recent manic episode	32.27 (7.43)	31.47 (5.93)	17 (56.7%)	15 (50%)	Middle 3 (10%) Secondary 12 (40%) University 15 (50%)	Middle 3 (10%) Secondary 12 (40%) University 15 (50%)	Not stated	Not stated	Antipsychotic n=30, mood stabiliser other than lithium n=28
Fakhry 2013 S2: recent depressive episode	31.60 (6.43)	As above	13 (43.3%)	As above	Middle 3 (10%) Secondary 15 (50%) University 12 (40%)	As above	Not stated	Not stated	Antipsychotic n=28, mood stabiliser other than lithium n=26, antidepressant n=22
Ibrahim 2009[a] & Normala 2010[a]	Mdn 37.5 IQR 20.0	Mdn 27.0 IQR 15.0	19 (47.5%)	10 (25%)	Not stated	Not stated	Onset Mdn 21 IQR 13 Duration 10.95 (9.04)	All types 3.83 (3.07)	Mood stabiliser only 27.5%, mood stabiliser + antipsychotic 55%, mood stabiliser + antidepressant 2.5%, antipsychotic only 15%
Juselius 2009[b]	44.2 (1.6)	47.8 (0.6)	15 (57.7%)	55 (48.2%)	Level attained[c] 4.1 (0.5)	Level attained[c] 4.1 (0.2)	Duration 20.5 (5.8)	Manic 3.8 (0.45)	Not stated
Osher 2011	41.3 (13.2)	53.7 (18.9)	25 (49%)	181 (37%)	12.8 (2.0)	15.3 (3.2)	Onset 24.0 (8.0)	Not stated	Mood stabiliser monotherapy n=12, neuroleptic monotherapy n=8, polytherapy n=31

BD, bipolar disorder; BD-I, bipolar disorder type I; HC, healthy comparison; M, mean; Mdn, median; IQR, interquartile range; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.

[a] Same sample.

[b] All participants were twins. BD sample (n=26) included 20 individuals whose co-twin did not have BD plus 3 pairs (6 individuals) concordant for BD. HC sample included n=114 twins (46 pairs + 22 individuals) with no history of BD in the participant or their co-twin.

[c] Level 4 is "vocation school or equivalent", after graduating high school.

Table S2 Characteristics of mixed BD samples

Author Year	% with BD-I	Age (years), M (SD)		Male gender, *n* (%)		Education (years), M (SD)		BD onset age and/or duration (years), M (SD)	BD past illness episodes, M (SD)	Psychotropic medication in BD sample
		BD	HC	BD	HC	BD	HC			
Barrera 2013	58%	48.21 (11.24)	46.04 (12.30)	12 (100%)	12 (100%)	12.33 (2.67)	12.50 (2.71)	Onset 25 (7.93)	Depressive 10.63 (13.5) Manic 5 (3.8)	Mood stabiliser 83.3%, antipsychotic 50%, anxiolytic 30%
Daban 2012	Not stated	41.12 (10.87)	46.53 (13.99)	23 (43.4%)	20 (33.3%)	14.21 (3.05)	12.49 (2.75)	Not stated	Not stated	Not stated
Martino 2014	51%	39.55 (10.83)	40.28 (12.03)	36 (36%)	12 (30%)	14.36 (2.36)	13.88 (2.77)	Onset 27.65 (9.49) Duration 11.18 (6.67)	Depressive 3.46 (2.01) Manic 3.18 (2.09)	Mood stabiliser 100%, antidepressant 38%, benzodiazepine 55%, antipsychotic 55%
Mur 2007	Not stated	42.6 (13.0)	42.2 (12.4)	22 (50%)	23 (50%)	10.5 (3.2)	12.5 (3.4)	Onset 25.6 (11.5) Duration 16.9 (11.67)	Manic 2.45 (2.5)	Lithium monotherapy n=20, lithium plus other n=24
Sánchez-Morla 2009	75%	43.5 (10.4)	43.8 (11.2)	30 (41.1%)	31 (46.3%)	12.5 (3.9)	14.1 (3.5)	Onset 26.2 (9.3) Duration 17.3 (10.5)	All types 13.3 (11.2) Manic 6.0 (5.6)	Lithium 39.8%, anticonvulsant 26.0%, lithium + anticonvulsant 30.1%, SSRI 30.1%, benzodiazepine 23.3%, typical antipsychotic 8.2%, atypical antipsychotic 26.0%
van der Werf-Eldering 2010	83%	Not stated[a]	40.8 (14.4)	Not stated[a]	27 (36%)	Not stated[a]	Level attained[b] 3.7 (1.1)	Not stated[a]	Not stated[a]	Not stated[a]

BD, bipolar disorder; HC, healthy comparison; M, mean; SD, standard deviation; SSRI, selective serotonin reuptake inhibitor.
[a] This information not reported in article for euthymic group separately.
[b] Ranging from 1 = primary school to 6 = PhD or higher degree obtained.

Table S3 Prevalence of cognitive impairment in BD-I versus BD-II samples, from Sparding et al. (2015)

Cognitive measure	Impairment prevalence[a,b]	
	BD-I[c]	BD-II[d]
D-KEFS TMT 2 number sequencing (executive/speed)	34%	19%
D-KEFS TMT 3 letter sequencing (executive/speed)	22%	14%
D-KEFS TMT 4 number-letter switching (executive)	48%	43%
D-KEFS Verbal Fluency 2 category (executive/language)	38%	39%
D-KEFS Verbal Fluency 3 category switching total (executive/language)	34%	27%
D-KEFS Verbal Fluency 3-v category switching accuracy (executive)	18%	11%
D-KEFS Color-Word 3 inhibition (executive)	29%	29%
D-KEFS Color-Word 4 inhibition/switching (executive)	27%	15%
D-KEFS Design Fluency 3 switching (executive)	32%	17%
D-KEFS Tower rule violations (executive)	34%	Not stated
WAIS-III Matrix Reasoning (abstract reasoning)	21%	30%
WAIS-III Similarities (abstract reasoning)	13%	11%
WAIS-III Letter-Number Sequencing (working memory)	22%	28%
WAIS-III Arithmetic (working memory)	37%	31%
WAIS-III Symbol Search (speed)	28%	19%
WAIS-III Digit-Symbol Coding (speed)	19%	11%
WAIS-III Digit Symbol Coding copy (speed)	35%	19%
WAIS-III Block Design (speed/visuospatial)	40%	44%
RCFT time to copy (speed/visuospatial)	23%	24%
RCFT 3-minute recall (visual memory)	23%	21%
RCFT recognition (visual memory)	22%	21%
WAIS-III Digit Symbol Coding free recall (visual memory)	25%	16%
WAIS-III Digit Symbol Coding pairing (visual memory)	26%	27%
WAIS-III Picture Completion (visuospatial)	14%	16%

BD-I, bipolar disorder type I; BD-II, bipolar disorder type II; D-KEFS, Delis-Kaplan Executive Function System; HC, healthy comparison; RCFT, Rey Complex Figure Test; SD, standard deviation; TMT, Trailmaking Test; WAIS-III, Wechsler Adult Intelligence Scale third edition.

[a] Impairment threshold is 1.25 SD from HC mean; HC impairment prevalence is 10.57% by definition. HC sample (n=86) was frequency matched to BD samples based on age, gender and years of education; details not reported in article.

[b] Standardised mean difference between BD and HC groups is not given in table because SD was not reported in article.

[c] BD-I (n=64) sample characteristics: age M=38 (SD=14); 48% male; education level M=3.7 (SD=1.1) (where 3=12 years, 4=13–15 years); age of onset M=19 (SD=9); past illness episodes (all types) M=19 (SD=26); medications: lithium 68%, antipsychotic 32%, antidepressant 31%, anticonvulsant 32%.

[d] BD-II (n=44) sample characteristics: age M=35 (SD=12); 45% male; education level M=3.9 (SD=1.2) (where 3=12 years, 4=13–15 years); age of onset M=18 (SD=11); past illness episodes (all types) M=18 (SD=18); medications: lithium 48%, antipsychotic 11%, antidepressant 41%, anticonvulsant 32%.

	1. Representative sample	2. Appropriate recruitment	3. Sample size	4. Subjects & setting described	5. Analysis coverage	6. Objective measure	7. Reliable measure	8. Statistical analysis	9. Confounding factors	10. Objective subgroupings
Altshuler 2004	N	Y	N	Y	?	Y	?	Y	Y	Y
Arslan 2014	N	Y	N	Y	?	Y	Y	Y	Y	Y
Barrera 2013	N	Y	N	N	?	Y	?	?	Y	NA
Cavanagh 2002	Y	Y	N	Y	Y	Y	Y	?	Y	NA
Cheung 2013	N	Y	Y	Y	?	Y	Y	Y	Y	Y
Daban 2012	Y	Y	Y	N	?	Y	?	Y	Y	Y
Doganavsargil-Baysal 2013	N	Y	N	N	?	Y	?	Y	Y	Y
Elshahawi 2011	Y	Y	Y	Y	?	Y	?	Y	Y	Y
Fakhry 2013	Y	Y	N	Y	?	Y	?	Y	Y	Y
Ferrier 1999	N	Y	N	Y	?	Y	?	Y	Y	Y
Frangou 2005	Y	Y	N	Y	?	Y	Y	Y	Y	NA
Goswami 2009	Y	Y	N	N	?	Y	Y	Y	Y	Y
Ibrahim[a] 2009	Y	Y	N	Y	?	Y	Y	?	Y	?
Jamrozinski 2009	Y	Y	N	N	?	Y	?	Y	Y	Y
Juselius[b] 2009	N	Y	N	Y	?	Y	?	Y	Y	NA
Kieseppä[b] 2005	N	Y	N	Y	Y	Y	?	Y	Y	Y
Lopera-Vásquez 2011	?	Y	N	N	?	Y	?	Y	Y	Y
López-Jaramillo 2010	N	Y	N	Y	?	Y	?	Y	Y	Y
Martino[c] 2008	N	Y	N	N	?	Y	Y	Y	Y	Y
Martino[c] 2011a	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2011b	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2011c	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Martino[c] 2014	N	Y	Y	Y	?	Y	Y	Y	Y	Y
Mur 2007	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Normala[a] 2010	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Osher 2011	Y	Y	Y	Y	?	Y	Y	Y	Y	Y

	1. Representative sample	2. Appropriate recruitment	3. Sample size	4. Subjects & setting described	5. Analysis coverage	6. Objective measure	7. Reliable measure	8. Statistical analysis	9. Confounding factors	10. Objective subgroupings
Pirkola[b] 2005	N	Y	N	Y	?	Y	Y	Y	Y	NA
Sánchez-Morla 2009	Y	Y	Y	Y	Y	Y	Y	Y	Y	NA
Sparding 2015	Y	Y	Y	Y	?	Y	?	Y	Y	Y
van der Werf-Eldering 2010	?	Y	N	Y	Y	Y	Y	Y	Y	Y

Figure S1 Ratings of risk of bias using Munn et al. critical appraisal tool for systematic reviews addressing questions of prevalence

[a] Studies contain overlapping samples.
[b] Studies contain overlapping samples.
[c] Studies contain overlapping samples.

Y	Yes
N	No
?	Unclear/unsure
NA	Not applicable

	1a. Design in abstract/title	1b. Abstract	2. Background	3. Objectives	4. Design	5. Setting	6a. Participants	7. Variables	8. Data sources/measurement	9. Bias	10. Study size	11. Quantitative variables	12a. Describe statistics	12b. Subgroups/interactions	12c. Missing data	12d. Account for sampling	12e. Sensitivity analyses	13a. Participant numbers at each stage[a]	13b. Non-participation reasons[a]	13c. Flow diagram	14a. Sample characteristics	14b. Number missing data	15. Outcome data	16a. Estimates & precision	16b. Category boundaries	16c. Absolute risk	17. Other analyses	18. Key results	19. Limitations	20. Interpretation	21. Generalisability	22. Funder and role
Altshuler 2004	N	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	N	Y	Y	Y	N
Arslan 2014	N	N	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	NA	N	NA	NA	Y	Y	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	N
Barrera 2013	N	N	Y	Y	N	N	N	Y	N	N	N	Y	N	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	NA	N	N	Y	N	N
Cavanagh 2002	Y	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	NA	N	N	Y	Y	Y	N	Y	N	Y	Y	NA	NA	Y	Y	Y	Y	Y	N
Cheung 2013	N	Y	Y	Y	N	N	N	Y	Y	N	N	N	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	Y	Y	Y
Daban 2012	N	Y	Y	Y	N	N	N	Y	Y	N	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	N	N	N
Doganavsargil-Baysal 2013	N	N	Y	Y	Y	N	N	N	Y	N	Y	Y	Y	Y	N	NA	NA	Y	Y	N	Y	N	Y	N	NA	NA	Y	Y	N	Y	N	Y
Elshahawi 2011	N	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	N	N	Y	Y
Fakhry 2013	Y	N	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	Y
Ferrier 1999	N	N	Y	Y	N	N	Y	Y	Y	N	N	N	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	NA	Y	N	Y	N	N
Frangou 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	NA	N	Y	NA	Y	Y	N	Y	N	Y	Y	NA	N	NA	Y	N	Y	N	N
Goswami 2009	N	N	Y	Y	N	N	Y	N	N	Y	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	N	NA	Y	Y	Y	Y	N	N
Ibrahim[b] 2009	Y	N	Y	Y	Y	N	N	Y	Y	N	N	N	N	Y	N	NA	NA	N	N	N	Y	N	Y	N	N	N	N	Y	Y	N	N	N
Jamrozinski 2009	Y	N	Y	Y	N	Y	N	Y	N	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	N	Y	N	N
Juselius[c] 2009	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	N	N	N	Y	N	Y	Y	Y	NA	Y	N	Y	Y	Y	N
Kieseppä[c] 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	NA	N	Y	N	Y	N	Y	Y	NA	NA	Y	N	Y	Y	Y	N
Lopera-Vásquez 2011	N	N	Y	Y	N	N	N	Y	N	N	N	Y	Y	NA	N	NA	NA	N	N	N	N	N	Y	N	NA	NA	NA	Y	N	N	N	N

	1a. Design in abstract/title	1b. Abstract	2. Background	3. Objectives	4. Design	5. Setting	6a. Participants	7. Variables	8. Data sources/measurement	9. Bias	10. Study size	11. Quantitative variables	12a. Describe statistics	12b. Subgroups/interactions	12c. Missing data	12d. Account for sampling	12e. Sensitivity analyses	13a. Participant numbers at each stage[a]	13b. Non-participation reasons[a]	13c. Flow diagram	14a. Sample characteristics	14b. Number missing data	15. Outcome data	16a. Estimates & precision	16b. Category boundaries	16c. Absolute risk	17. Other analyses	18. Key results	19. Limitations	20. Interpretation	21. Generalisability	22. Funder and role
López-Jaramillo 2010	N	N	Y	Y	Y	N	N	Y	N	Y	N	Y	Y	NA	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	NA	Y	N	Y	N	N
Martino[d] 2008	N	Y	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	N	Y	Y	Y	N
Martino[d] 2011a	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	Y	Y	N	N
Martino[d] 2011b	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	N	Y	Y	N	N
Martino[d] 2011c	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	Y	Y	Y	N
Martino[d] 2014	N	Y	Y	Y	N	N	N	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	Y	Y	N	N
Mur 2007	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	NA	NA	Y	Y	Y	Y	N	Y	N	NA	NA	Y	Y	N	Y	Y	N
Normala[b] 2010	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	Y	Y	Y	N
Osher 2011	N	N	Y	Y	N	Y	N	Y	Y	Y	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	NA	NA	Y	Y	N	N	N	N
Pirkola[c] 2005	N	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	NA	N	Y	Y	N	N	N	Y	N	Y	Y	Y	NA	NA	Y	Y	Y	Y	N
Sánchez-Morla 2009	N	N	Y	Y	N	N	N	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	N	Y	NA	Y	Y	N	Y	Y	N
Sparding 2015	N	N	Y	Y	N	N	Y	Y	Y	N	N	Y	Y	Y	N	NA	NA	N	N	N	Y	N	Y	Y	NA	NA	Y	Y	Y	N	Y	Y
van der Werf-Eldering 2010	N	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	NA	Y	Y	Y	N	Y	N	Y	Y	Y	NA	Y	Y	N	Y	Y	Y

Figure S2 Ratings of reporting bias using STROBE checklist for cross-sectional studies

[a] Rated Yes if information reported for bipolar disorder sample, at a minimum.
[b] Studies contain overlapping samples.
[c] Studies contain overlapping samples.
[d] Studies contain overlapping samples.

Y Yes
N No
NA Not applicable

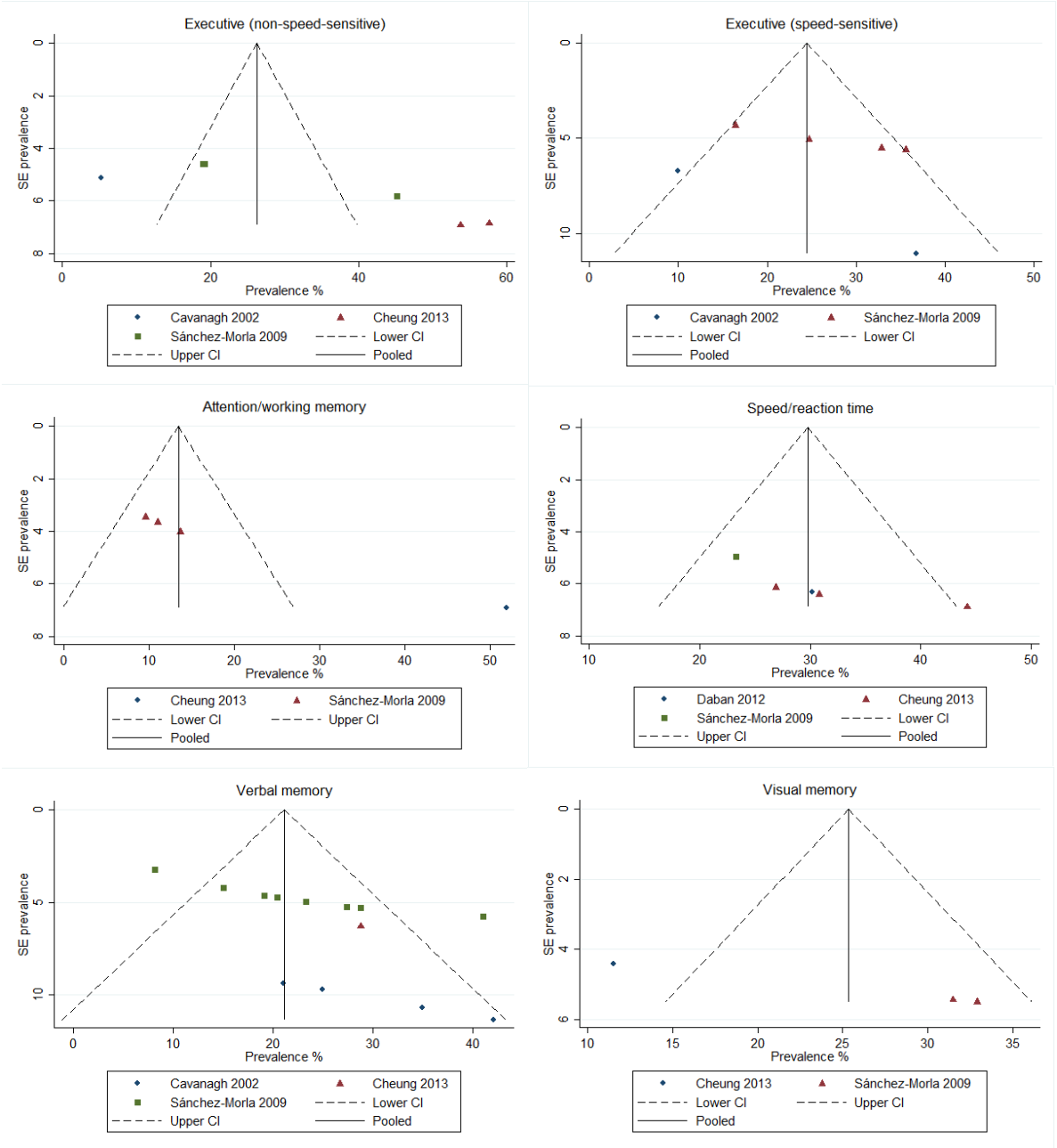


Figure S3 Funnel plots with pseudo 95% confidence limits

SE, standard error.
Plots show cognitive impairment prevalence estimates and standard errors by cognitive domain, using results reported at the 5th percentile impairment threshold. Standard errors were calculated as follows: sqrt((prevalence*(100-prevalence))/n). Some studies contributed more than one data point to the plot; data points are labelled by study.

	Age ↑	Gender	Education ↓	Premorbid ability ↓	Ethnicity	Parental socioeconomic status ↓	Age at BD onset ↓	Duration of illness ↑	Residual affective symptoms ↑	Duration of euthymia ↓	Bipolar disorder subtype (BD-I)	No. of hospitalisations ↑	No. of episodes (all types) ↑	No. of episodes (manic) ↑	No. of episodes (depressed) ↑	History of psychotic symptoms	Any psychotropic medication	Mood stabiliser	Antipsychotic	Antidepressant	Benzodiazepine	Electroconvulsive therapy	Inflammatory biomarkers ↑	Alcohol/substance use disorder	Comorbid ADHD	Obstetric complications	Family history psychiatric illness
Altshuler 2004																											
Arslan 2014																											
Barrera 2013																											
Cavanagh 2002																											
Cheung 2013																											
Daban 2012		M<F																									
Doganavsargil-Baysal 2013																											
Elshahawi 2011		M<F																									
Ferrier 1999													a														
Frangou 2005																											
Goswami 2009																											
Ibrahim[b] 2009																											
Jamrozinski 2009																		c									
Juselius[d] 2009																c											
Kieseppä[d] 2005																											
Lopera-Vásquez 2011																											

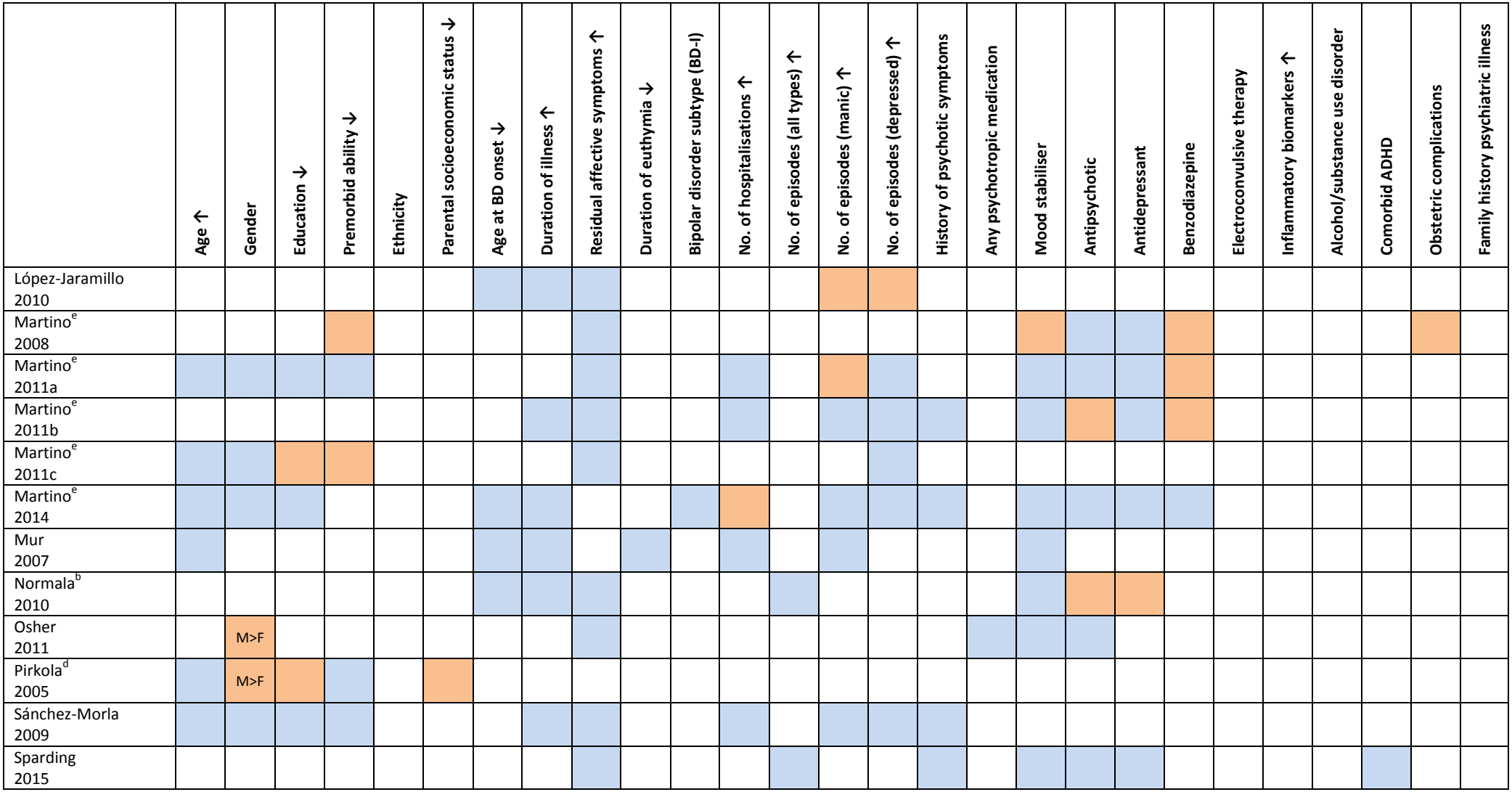


Figure S4 Variables associated with cognitive impairment

ADHD, attention deficit hyperactivity disorder; BD, bipolar disorder; F, female; M, male.
[a] Analysis based on number of recent episodes as well as time to recovery.
[b] Studies contain overlapping samples.
[c] Positive association with cognitive performance on some scores.
[d] Studies contain overlapping samples.
[e] Studies contain overlapping samples.